

# Higher-order interactions stabilize dynamics in competitive network models

Jacopo Grilli, György Barabás, Matthew J. Michalska-Smith & Stefano Allesina

Nature (2017), 548:210-213

## Abstract

Ecologists have long sought a way to explain how the staggering biodiversity observed in nature is maintained. On the one hand, simple models of interacting competitors cannot produce the stable persistence of very large ecological communities<sup>1-5</sup>; on the other hand, neutral models<sup>6-9</sup>, in which species do not interact and diversity is maintained by immigration and speciation, yield unrealistically small fluctuations in population abundance<sup>10</sup>, and a strong positive correlation between a species' abundance and its age<sup>11</sup>, contrary to empirical evidence. Models allowing for the robust persistence of large communities of interacting competitors are lacking. Here we show that very rich communities could persist thanks to the stabilizing role of higher-order interactions<sup>12,13</sup>, in which the presence of a species influences the interaction between other species. The existence of higher-order interactions has been debated in ecology for decades<sup>14-16</sup>, but their role in shaping ecological communities is still understudied<sup>5</sup>. Our results show that higher-order interactions can have dramatic effects on the dynamics of ecological systems, producing models in which coexistence is robust to the perturbation of both population abundances and parameter values. Introducing higher-order interactions has strong effects on models of closed ecological communities, as well as simulations of open communities in which new species are constantly introduced. Notably, in our framework higher-order interactions are completely defined by pairwise interactions, easing empirical parameterization and validation of our models.

Here we study deterministic models describing communities in which the number of individuals is large and the system is isolated (e.g., bacterial strain competition in laboratory conditions<sup>17</sup>); in the Supplementary Information (S4) we examine the case in which the dynamics are stochastic, which best describe communities in which the number of individuals is finite. Finally, we allow new species to be introduced at a given rate, allowing for a comparison with neutral models (Supplementary Information S5).

While our results hold for a wide class of systems, to exemplify our findings we consider the dynamics of a forest in which there is a fixed, large number of trees, so that we can simply track  $x_i(t)$ , the proportion of trees of species  $i$  at time  $t$ , with  $\sum_i x_i(t) = 1$ . At each step, a randomly selected tree dies, opening a gap in the canopy (i.e., we initially assume identical per capita death rates for all species). This event ignites competition among seedlings to fill the gap. Suppose that all individuals produce the same number of seedlings, and that we pick two seedlings at random, with the winner of the competition filling the gap (Fig. S1). The matrix  $\mathbf{H}$  encodes the dominance relationships among the species:  $H_{ij}$  is the probability that the first seedling, belonging to species  $i$ , wins against the second seedling, belonging to species  $j$ . Clearly,  $H_{ii} = 1/2$  for all  $i$ , and  $H_{ij} + H_{ji} = 1$  for all

$i$  and  $j$ . If all  $H_{ij} = 1/2$ , we recover a neutral model. At the other extreme lies a model in which each pair  $(H_{ij}, H_{ji})$  is either  $(1, 0)$  or  $(0, 1)$  (i.e., species  $i$  always wins or always loses against  $j$ ), in which case  $\mathbf{H}$  is called a “tournament matrix”<sup>18</sup>. A number of results have been derived for this case<sup>19</sup>, showing that coexistence is possible when species form “intransitive cycles” of competitive dominance, such as in the rock-paper-scissors game<sup>20</sup>. Here we extend these previous findings<sup>19</sup> to the most general case in which interactions range from neutral to complete dominance.

We can approximate the dynamics of the  $n$  species as

$$\frac{dx_i(t)}{dt} = x_i(t) \left( 2 \sum_j H_{ij} x_j(t) - 1 \right), \quad (1)$$

where  $-x_i(t)$  models the death process, and  $x_i(t)2\sum_j H_{ij}x_j(t)$  is the probability of picking two seedlings of species  $i$  and  $j$ , with  $i$  winning the competition. The factor 2 arises from the fact that we could pick  $i$  first and  $j$  second, or vice versa, with the same outcome.

Simple manipulations (Supplementary Information S1) show that these equations are equivalent to the system

$$\frac{dx_i(t)}{dt} = x_i(t) \sum_j P_{ij} x_j(t), \quad (2)$$

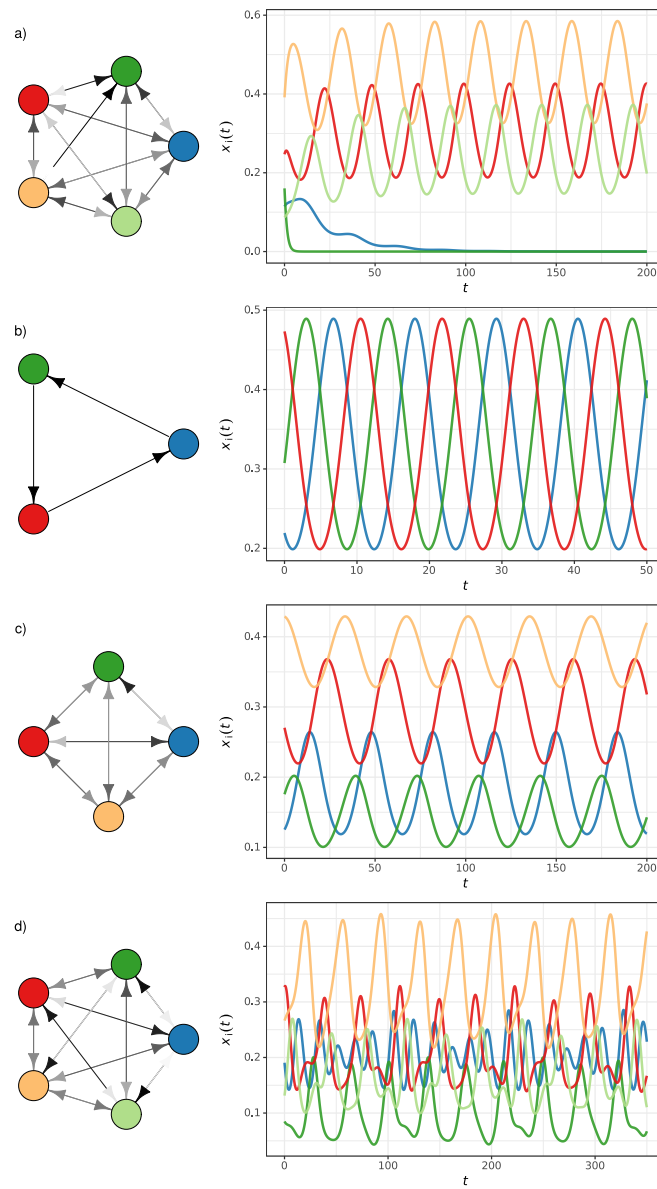
which is the celebrated replicator equation<sup>21,22</sup> for a zero-sum, symmetric matrix game with two players and payoffs encoded in the skew-symmetric matrix  $\mathbf{P} = \mathbf{H} - \mathbf{H}^t$ . This equation is at the core of evolutionary game theory, with applications spanning multiple fields<sup>23,24</sup>.

Thanks to this equivalence, we are able to characterize the dynamics. Unless specified, we assume  $\mathbf{H}$  to be of full rank, i.e., all of its eigenvalues are nonzero. We show in the Supplementary Information that violations of this assumption are unbiological, amounting to degenerate cases in which slightly altering the parameters dramatically changes the outcome. Suppose that we start with  $n$  species and initial conditions  $x_i(0) > 0$ , and that we let the dynamics unfold. Once the transient dynamics have elapsed, we find  $k \leq n$  coexisting species, with  $k$  being odd. The  $n - k$  species that go extinct do so irrespective of initial conditions, and the  $k$  coexisting species cycle neutrally around a unique equilibrium point  $x^*$  (Fig. 1, Supplementary Information S1).

How large is  $k$  when we build the matrix  $\mathbf{H}$  at random? When drawing  $H_{ij}$  (with  $i < j$ ) from the uniform distribution  $\mathcal{U}[0, 1]$  and setting the corresponding  $H_{ji} = 1 - H_{ij}$ , we find that the probability of having  $k$  species coexisting when starting with  $n$ ,  $p(k|n) = 0$  when  $k$  is even, and  $p(k|n) = \binom{n}{k} 2^{1-n}$  when  $k$  is odd<sup>25</sup> (Fig. S2). This matches what found for tournament games<sup>18,19,26</sup>, in which dominance is complete: we expect half of the initial species to coexist, irrespective of the choice of  $n$ ; moreover, monodominance is extremely rare, and about as rare as the coexistence of all species. Thus, this theory generates high biodiversity without the need to fine-tune parameters.

This model can generate any species-abundance distribution: for any choice of  $x^*$ , we can build infinitely many matrices  $\mathbf{H}$  such that Eqs. 1 and 2 have  $x^*$  as an equilibrium (Supplementary Information S1). Note that this is true irrespective of the fact that  $x^*$  contains an even or odd number of species (Fig. 1)—but the case in which an even number of species coexist is degenerate: the system has infinitely many neutrally stable equilibria, and a slight change of  $\mathbf{H}$  would result in the extinction of at least one species (Supplementary Information, Fig. S4).

In summary, the model in Eq. 1 can lead to arbitrarily many species coexisting even when competitive abilities are drawn at random; moreover, it can generate any possible species-abundance distribution. While the neutral cycling around equilibrium is problematic (such cycles are not observed in nature, and would lead to monodominance in a noisy, stochastic world, Supplementary



**Figure 1: Sampling two seedlings: cycles.** Dynamics of a forest where two randomly sampled seedlings compete to fill the gap in the canopy opened by the death of a tree. We represent the competitive relationships between the species as a graph in which each colored node corresponds to a species, and nodes  $i$  and  $j$  are connected by two edges whose color shade stands for the probability that  $i$  wins against  $j$  ( $H_{ij}$ , darker arrowheads correspond to higher probability), and that  $j$  wins against  $i$  ( $H_{ji}$ ), respectively. For all pairs,  $H_{ij} + H_{ji} = 1$ . For each system, we show a time series obtained by integrating the dynamics of Eq. 1. The y-axis represents  $x_i(t)$ , the relative abundance of species  $i$  at time  $t$ . a) When starting with  $n$  species,  $n - k$  species go extinct, and  $k$  coexist. Given a matrix  $\mathbf{H}$ , the identity of the species coexisting or going extinct is the same irrespective of initial conditions. The  $k$  species that coexist cycle neutrally around a single equilibrium point. b) The same is found when dominance is complete, such as in the rock-paper-scissors game<sup>19</sup>. c) For any possible species-abundance distribution  $x^*$ , we can build a matrix  $\mathbf{H}$  such that the species coexist and  $x^*$  is an equilibrium of Eq. 1 (Supplementary Information S1). This is true even when  $x^*$  contains an even number of species—though this case is not robust to small changes in parameters (Supplementary Information, Fig. S4). d) The same holds for any number of species in the system.

Information S4), the main issue with this model is that it is highly unrobust: any deviation from perfectly identical death rates and fecundities for all species destabilizes dynamics, leading to monodominance (Supplementary Information, Fig. S4).

Following recent mathematical results<sup>27</sup>, we explore a possible solution to this problem. So far, we have taken exactly two seedlings, competing with each other to fill the gap in the canopy. In nature, we would observe a much richer seedbank, potentially leading to the competition among many seedlings. We therefore study a model in which we take three seedlings at random, compete the first with the second, and the winner with the third. The deterministic approximation of this model reads

$$\frac{dx_i(t)}{dt} = x_i(t) \left( \sum_{j,k} (2H_{ij}H_{ik} + H_{ij}H_{jk} + H_{ik}H_{kj}) x_j(t)x_k(t) - 1 \right), \quad (3)$$

where  $H_{ij}H_{ik}$  is the probability that  $i$  beats both  $j$  and  $k$ ,  $H_{ij}H_{jk}$  that  $j$  beats  $k$ , but ultimately is beaten by  $i$ , and  $H_{ik}H_{kj}$  that first  $k$  beats  $j$ , and then  $i$  beats  $k$ . Surprisingly, this small modification leads to a major change in the dynamics: though the equilibrium point is unchanged, it is now globally stable (Fig. 2 and Supplementary Information S1). Increasing the number of seedlings that compete to fill each gap simply accelerates the dynamics, speeding up convergence to the equilibrium (Fig. S3).

While the model in which we sample two seedlings yields the replicator equation for a two-player, symmetric matrix game, Eq. 3 is equivalent to the replicator equation for a three-player game (Supplementary Information S1):

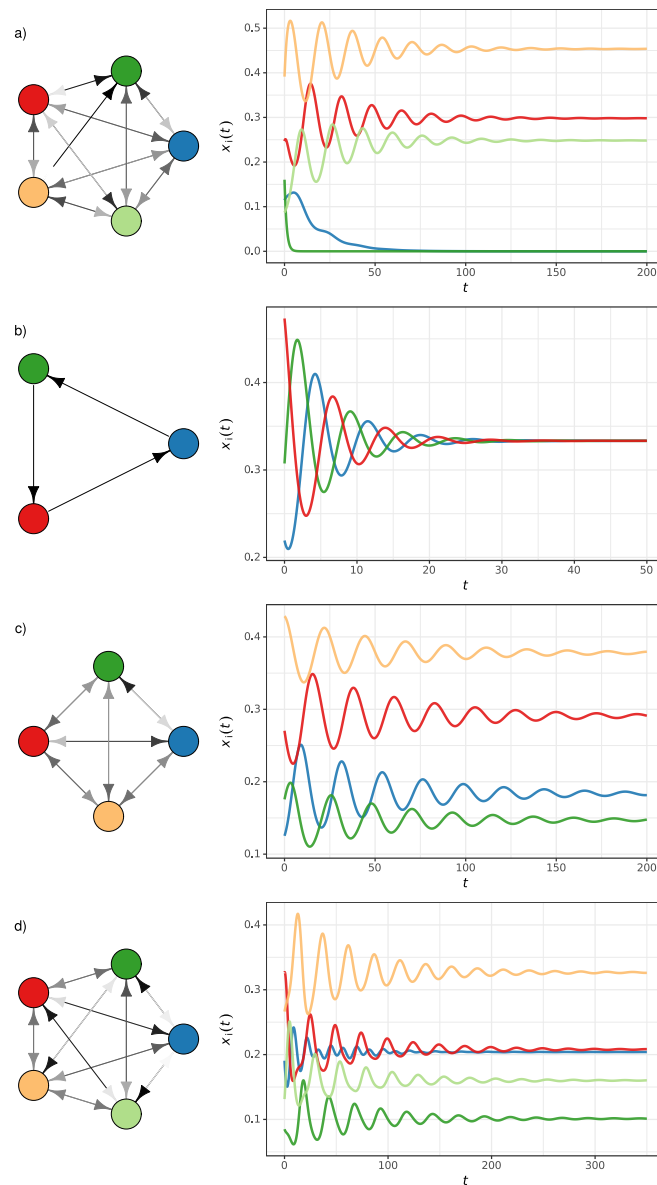
$$\frac{dx_i(t)}{dt} = x_i(t) \sum_{j,k} P_{ijk} x_j(t)x_k(t), \quad (4)$$

where  $\mathbf{P}$  is a 3-index tensor encoding the payoff of player 1 playing strategy  $i$  when player 2 plays  $j$  and player 3 plays  $k$ . The payoffs can be calculated from the matrix  $\mathbf{H}$ :  $P_{ijk} = 2H_{ij}H_{ik} - H_{ji}H_{jk} - H_{ki}H_{kj}$  (where the first term includes the probability of  $i$  winning against both  $j$  and  $k$ , and the remaining two terms the probability that either  $j$  or  $k$  dominate).

This latter formulation makes it clear that the stabilizing effect is due to higher-order interactions<sup>5,12</sup>: suppose the matrix  $\mathbf{H}$  is constructed as in a rock-paper-scissors game; then the presence of the rock-plant can reverse the outcome of the competition between the paper- and the scissors-plant. In our model, higher-order interactions do not alter equilibrium values, but have a dramatic stabilizing effect, leading to globally stable fixed points instead of neutral cycles. Including fourth- or higher-order terms simply accelerates the convergence to equilibrium. As such, as long as there is a chance of competing more than two seedlings at a time, dynamics will converge. Most importantly, results are qualitatively robust to the perturbations of the death rates and fecundities of the competitors (Supplementary Information, Fig. S4).

One formidable challenge of estimating higher-order interactions empirically is that for  $n$  species we have  $\binom{n}{2} = n(n-1)/2$  pairs of interactions, but the number of triplets is much higher  $\binom{n}{3} = n(n-1)(n-2)/6$ —requiring many experiments. Instead of introducing new coefficients, here we have chosen the most “natural” and conservative parameterization: higher-order interactions are fully determined by pairwise interactions, as shown by the fact that we can write all models in terms of the pairwise relationships encoded in  $\mathbf{H}$ . This makes the models empirically testable by, for example, competing bacteria in laboratory conditions<sup>17</sup>.

We have shown the equivalence of models in which competition happens in a sequence of bouts (Eq. 3) with models in which interactions are simultaneous and involve more than two species at a time (Eq. 4). Because of a separation of timescales (the filling of a gap is fast, compared



**Figure 2: Sampling three seedlings: stability.** When we sample three seedlings at a time instead of two, and we compete the first with the second and the winner with the third, the equilibrium point is unchanged, but is now globally attractive. The four cases correspond to those in Fig. 1.

to the lifespan of trees), the two types of models have the same deterministic form, blurring the traditionally-held distinction between so-called interaction chains and “proper” higher-order interactions<sup>5,28</sup>. Our results may have important implications for a variety of ecological systems; for example, in models in which reproduction is not instantaneously coupled with consumption, an animal could consume a resource, but be consumed before reproduction—yielding the same mechanism that stabilizes our competitive communities when we sample three seedlings at a time. Similarly, the stabilizing role of higher-order interactions in random replicator equations has been recently proposed<sup>29</sup>, and our analytical results shed light on these findings.

Moving from deterministic to stochastic models, we find that the presence of higher-order interactions, which make equilibrium points attractive, dramatically increase<sup>30</sup> time to extinction in isolated systems, allowing for the prolonged coexistence of species (Supplementary Information S4). When we open the system to the introduction of new species (Supplementary Information S5), we recover many of the main results of neutral theory, but remove the artifactual relationship between a species age and its abundance—one of the main drawbacks of neutral models<sup>11</sup>.

Our results strengthen the theory of coexistence in zero-sum competitive networks in several ways. First, we have widespread coexistence without having to invoke either of two extreme cases: perfect ecological equivalence (neutral model) or complete dominance (coexistence through intransitive competition). In nature, the outcome of competition could be mediated by a number of factors (e.g., soil chemistry, presence of consumers), so that competitive dominance could range from neutral to complete. Second, many species coexist even when we draw parameters at random, meaning that the results are highly robust. Third, in this formulation, the notion of intransitivity, which is central to coexistence in competitive networks in which dominance is complete<sup>19</sup>, is no longer necessary for coexistence (Supplementary Information S3). Fourth, the artifact of neutral cycling is due to the choice of only two competitors per bout—a choice dictated by mathematical convenience rather than by empirical evidence. Including more biological realism in the form of multiple competing species removes the artifact, leading to dynamics that are stable against perturbations of species abundances and robust against changing model parameters.

## Acknowledgements

Thanks to J. Aljadeff, F. Brandl, J.A. Capitán, J-F. Laslier, J.M. Levine, C.A. Marcelo Serván, and E. Sander for comments and discussions. E. Leigh and four anonymous referees provided constructive feedback. J.G. supported by the Human Frontier Science Program; S.A. and G.B. by NSF DEB-1148867; M.J. M-S. by US Department of Education grant P200A150101.

## References

- [1] May, R. M. Will a large complex system be stable? *Nature* **238**, 413–414 (1972).
- [2] Clark, J. S. *et al.* High-dimensional coexistence based on individual variation: a synthesis of evidence. *Ecological Monographs* **80**, 569–608 (2010).
- [3] Barabás, G., J. Michalska-Smith, M. & Allesina, S. The effect of intra-and interspecific competition on coexistence in multispecies communities. *The American Naturalist* **188**, E1–E12 (2016).
- [4] D’Andrea, R. & Ostling, A. Challenges in linking trait patterns to niche differentiation. *Oikos* **125**, 1369–1385 (2016).
- [5] Levine, J., Bascompte, J., Adler, P. & Allesina, S. Beyond pairwise coexistence: biodiversity maintenance in complex ecological communities. *Nature* (**in press**) (2017).
- [6] Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*, vol. 32 (Princeton University Press, 2001).
- [7] Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037 (2003).

- [8] Alonso, D., Etienne, R. S. & McKane, A. J. The merits of neutral theory. *Trends in Ecology & Evolution* **21**, 451–457 (2006).
- [9] Azade, S. *et al.* Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics* **88**, 035003 (2016).
- [10] Chisholm, R. A. *et al.* Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecology Letters* **17**, 855–865 (2014).
- [11] Chisholm, R. A. & O'Dwyer, J. P. Species ages in neutral biodiversity models. *Theoretical Population Biology* **93**, 85–94 (2014).
- [12] Billick, I. & Case, T. J. Higher order interactions in ecological communities: what are they and how can they be detected? *Ecology* **75**, 1529–1543 (1994).
- [13] Werner, E. E. & Peacor, S. D. A review of trait-mediated indirect interactions in ecological communities. *Ecology* **84**, 1083–1100 (2003).
- [14] Case, T. J. & Bender, E. A. Testing for higher order interactions. *The American Naturalist* **118**, 920–929 (1981).
- [15] Abrams, P. A. Arguments in favor of higher order interactions. *The American Naturalist* **121**, 887–891 (1983).
- [16] Kareiva, P. Special feature: Higher Order Interactions as a Foil to Reductionist Ecology. *Ecology* **75**, 1527–1528 (1994).
- [17] Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1**, 0109 (2017).
- [18] Fisher, D. C. & Ryan, J. Optimal strategies for a generalized “scissors, paper, and stone” game. *American Mathematical Monthly* **99**, 935–942 (1992).
- [19] Allesina, S. & Levine, J. M. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences USA* **108**, 5638–5642 (2011).
- [20] Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418**, 171–174 (2002).
- [21] Taylor, P. D. & Jonker, L. B. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences* **40**, 145–156 (1978).
- [22] Hofbauer, J., Schuster, P. & Sigmund, K. A note on evolutionary stable strategies and game dynamics. *Journal of Theoretical Biology* **81**, 609–612 (1979).
- [23] Hofbauer, J. & Sigmund, K. Evolutionary game dynamics. *Bulletin of the American Mathematical Society* **40**, 479–519 (2003).
- [24] Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. *Science* **303**, 793–799 (2004).
- [25] Brandl, F. The distribution of optimal strategies in symmetric zero-sum games. *arXiv preprint arXiv:1611.06845* (2016).

- [26] Fisher, D. C. & Reeves, R. B. Optimal strategies for random tournament games. *Linear Algebra and its Applications* **217**, 83–85 (1995).
- [27] Laslier, B. & Laslier, J.-F. Reinforcement learning from comparisons: Three alternatives is enough, two is not. *arXiv preprint arXiv:1301.5734* (2013).
- [28] Wootton, J. T. Indirect effects and habitat use in an intertidal community: interaction chains and interaction modifications. *The American Naturalist* 71–89 (1993).
- [29] Bairey, E., Kelsic, E. D. & Kishony, R. High-order species interactions shape ecosystem diversity. *Nature Communications* **7** (2016).
- [30] Reichenbach, T., Mobilia, M. & Frey, E. Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model. *Physical Review E* **74**, 051907 (2006).



## Methods

To exemplify the role of higher-order interactions in shaping ecological dynamics, we consider a model of a forest in which whenever a tree dies, a certain number of seedlings compete to fill the gap in the canopy (Fig. S1). We start by writing the microscopic rate  $W_{ij}^{(m)}$  at which species  $j$  loses an individual ( $\eta_j \rightarrow \eta_j - 1$ ), while species  $i$  gains an individual ( $\eta_i \rightarrow \eta_i + 1$ ). The index  $m$  specifies that we consider the case in which  $m$  seedlings at a time compete to fill in the gap. When there are many individuals we can track proportions ( $x_i(t) = \eta_i(t) / \sum_j \eta_j(t)$ ), and setting  $m = 2$ , we can write this rate as:

$$W_{ij}^{(2)} = d_j x_j \frac{f_i x_i}{\sum_l f_l x_l} \sum_k 2H_{ik} \frac{f_k x_k}{\sum_l f_l x_l} \quad (5)$$

in which  $f_i$  and  $d_i$  are the fecundity and death rate of trees belonging to species  $i$ . Using the notation  $F(x) = \sum_l f_l x_l$ , the term  $f_i x_i / \sum_l f_l x_l = f_i x_i / F(x)$  is the proportion of seeds in the seedbank belonging to species  $i$ . Finally, the matrix  $\mathbf{H}$  encodes the probability of winning for every pair of species, so that  $H_{ik}$  is the probability of seedling of species  $i$  beating those of species  $k$ , and the factor 2 arises from the fact that we could sample  $i$  first and  $k$  second, or vice versa. Then, the term  $W_{ij}^{(2)}$  can be interpreted as the rate at which a) a tree of species  $j$  dies, and b) two seedlings of species  $i$  and  $k$  are sampled, with  $i$  filling the gap. Because the identity of  $k$  does not matter, we sum over all possible choices.

When the number of trees is sufficiently large, we can neglect stochasticity and write:

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(2)} - W_{ji}^{(2)} \right) = x_i \left( \frac{D(x)}{F(x)^2} \sum_k f_i H_{ik} f_k x_k - d_i \right) \quad (6)$$

where we have introduced  $D(x) = \sum_l d_l x_l$ . In Supplementary Information S4 we present stochastic simulations that show strong agreement with the predictions of this deterministic approximation. We can derive equations like Eq. 6 for any choice of  $m$ . We write:

$$W_{ij}^{(m)} = d_j x_j q_i^{(m)} \quad (7)$$

where  $q_i^{(m)}$  is the probability that a seedling of species  $i$  wins when competing against  $m - 1$  other seedlings. We build  $q_i^{(m)}$  recursively:

$$\begin{cases} q_i^{(1)} = \frac{f_i x_i}{F(x)} \\ \dots \\ q_i^{(m)} = \frac{f_i x_i}{F(x)} \sum_k H_{ik} q_k^{(m-1)} + q_i^{(m-1)} \sum_k H_{ik} \frac{f_k x_k}{F(x)} \end{cases} \quad (8)$$

which has a simple interpretation:  $\frac{f_i x_i}{F(x)} \sum_k H_{ik} q_k^{(m-1)}$  is the probability that  $i$  wins against the winner of the competition of involving the first  $m - 1$  seedlings, while  $q_i^{(m-1)} \sum_k H_{ik} \frac{f_k x_k}{F(x)}$  is the probability that  $i$  is the winner of the competition among the first  $m - 1$  seedlings, and beats the last seedling. Consistently with the fact that these are probabilities,  $\sum_i q_i^{(m)} = 1$  for any  $m$ . In this way, we can for example recover the rate  $W_{ij}^{(2)}$  we have introduced above:

$$W_{ij}^{(2)} = d_j x_j q_i^{(2)} = d_j x_j \left( \frac{f_i x_i}{F(x)} \sum_k H_{ik} q_k^{(1)} + q_i^{(1)} \sum_k H_{ik} \frac{f_k x_k}{F(x)} \right) = d_j x_j \frac{f_i x_i}{F(x)} \sum_k 2H_{ik} \frac{f_k x_k}{F(x)} \quad (9)$$

In general, the dynamics when considering  $m$  seedlings become:

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(m)} - W_{ji}^{(m)} \right) = D(x) q_i^{(m)} - d_i x_i. \quad (10)$$

Using this formulation, we write the system of equations describing the model in which three seedlings are sampled. We calculate  $q_i^{(3)}$ :

$$\begin{aligned} q_i^{(3)} &= \frac{f_i x_i}{F(x)} \sum_k H_{ik} q_k^{(2)} + q_i^{(2)} \sum_k H_{ik} \frac{f_k x_k}{F(x)} \\ &= \frac{f_i x_i}{F(x)} \sum_{j,k} 2H_{ij} \frac{f_j x_j}{F(x)} H_{jk} \frac{f_k x_k}{F(x)} + \frac{f_i x_i}{F(x)} \sum_{j,k} 2H_{ij} \frac{f_j x_j}{F(x)} H_{ik} \frac{f_k x_k}{F(x)} \\ &= \frac{f_i x_i}{F(x)} \sum_{j,k} \left( (2H_{ij} H_{jk} + 2H_{ij} H_{ik}) \frac{f_j x_j}{F(x)} \frac{f_k x_k}{F(x)} \right) \\ &= \frac{f_i x_i}{F(x)} \sum_{j,k} \left( (H_{ij} H_{jk} + H_{ik} H_{kj} + 2H_{ij} H_{ik}) \frac{f_j x_j}{F(x)} \frac{f_k x_k}{F(x)} \right) \end{aligned}$$

yielding the system of equations

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(3)} - W_{ji}^{(3)} \right) = x_i \left( \frac{D(x)}{F(x)^3} f_i \sum_{j,k} (2H_{ij} H_{ik} + H_{ij} H_{jk} + H_{ik} H_{kj}) f_j x_j f_k x_k - d_i \right) \quad (11)$$

Having derived this general case, we dedicate Supplementary Information S1 to the study of the simplified model we have introduced in the main text, in which  $f_i = d_i = 1$  for all  $i$ . It is easy to derive a number of results for this simplified formulation, including the existence and uniqueness of the coexistence equilibrium, the stability properties of such equilibrium when sampling two or more than two seedlings, the expected number of coexisting species when  $\mathbf{H}$  is random, and the construction of an algorithm that takes as input a desired species-abundance distribution  $x^*$ , and produces infinitely many  $\mathbf{H}$  such that  $x^*$  is an equilibrium of the system.

In Supplementary Information S2, we return to the more general case introduced above to test the robustness of our findings when we relax the strong constraint of identical physiological rates for all species. Supplementary Information S3 is dedicated to the discussion of intransitivity. Finally, Supplementary Information S4 and S5 extend these models to situations in which the number of individuals is finite, and therefore demographic stochasticity becomes important. We first consider the case of an isolated ecological community (S4), and then open the community to the introduction of new species by immigration or speciation (S5), allowing for a direct contrast with neutral models.

# Higher-order interactions stabilize dynamics in competitive network models

## Supplementary Information

Jacopo Grilli, György Barabás, Matthew J. Michalska-Smith & Stefano Allesina

Nature (2017), 548:210-213

### S1 Identical physiological rates

In the Methods we have shown that, when considering systems with a large number of individuals, our model (depicted in Fig. S1) can be written as a system of differential equations:

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(m)} - W_{ji}^{(m)} \right) = D(x) q_i^{(m)} - d_i x_i, \quad (\text{S1})$$

where  $m$  is the number of seedlings that compete for filling each gap,  $W_{ij}^{(m)}$  is the rate at which individuals of species  $j$  are replaced by those of species  $i$ , and the function  $q_i^{(m)}$  can be built recursively:

$$\begin{cases} q_i^{(1)} = \frac{f_i x_i}{F(x)} \\ \dots \\ q_i^{(m)} = \frac{f_i x_i}{F(x)} \sum_k H_{ik} q_k^{(m-1)} + q_i^{(m-1)} \sum_k H_{ik} \frac{f_k x_k}{F(x)} \end{cases}. \quad (\text{S2})$$

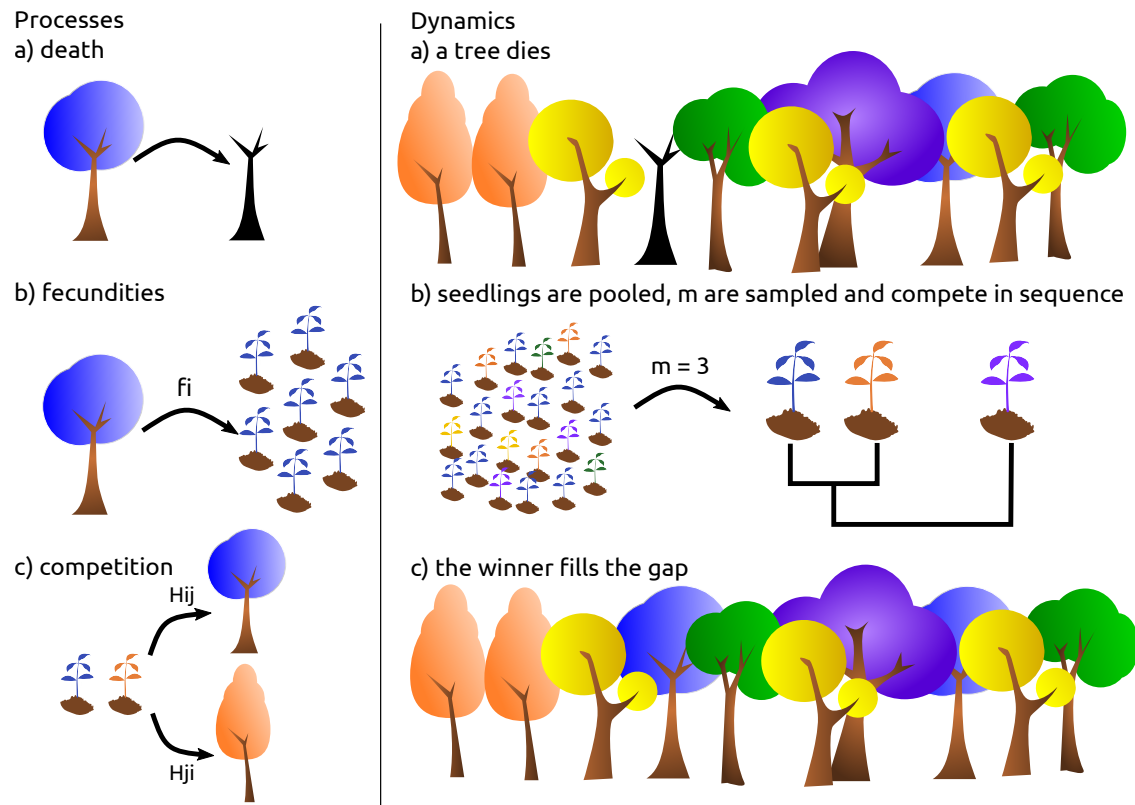
For the cases of two seedlings competing, the equations describing the dynamics become:

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(2)} - W_{ji}^{(2)} \right) = x_i \left( \frac{D(x)}{F(x)^2} \sum_k f_i H_{ik} f_k x_k - d_i \right), \quad (\text{S3})$$

and those describing the case in which we sample three seedlings become:

$$\frac{dx_i}{dt} = \sum_j \left( W_{ij}^{(3)} - W_{ji}^{(3)} \right) = x_i \left( \frac{D(x)}{F(x)^3} f_i \sum_{j,k} (2H_{ij}H_{ik} + H_{ij}H_{jk} + H_{ik}H_{kj}) f_j x_j f_k x_k - d_i \right). \quad (\text{S4})$$

In this section, we study these models when we make the simplifying assumption of equal physiological rates ( $d_i = f_i = 1$ ) for all species.



**Figure S1: Illustration of the model**, Left: the dynamics are controlled by the following parameters: (a)  $d_i$ , the death rate of species  $i$ ; (b)  $f_i$ , the fecundity of species  $i$ ; and (c) the matrix  $\mathbf{H}$ , detailing the probability that seedling from one species win against seedling of other species when competing to fill a gap. Right: a) whenever a tree dies, we pool the seedlings, and select  $m$  to compete (b), with the winner filling the gap (c).

### S1.1 Sampling two seedlings

When setting all  $f_i = d_i = 1$  in Eq. S3, we obtain Eq. 1 of the main text:

$$\frac{dx_i}{dt} = x_i 2 \sum_j H_{ij} x_j - x_i \quad (\text{S5})$$

We want to show that this is equivalent to the replicator equation, Eq. 2 of the main text:

$$\frac{dx_i}{dt} = x_i \sum_j P_{ij} x_j \quad (\text{S6})$$

where  $\mathbf{P} = \mathbf{H} - \mathbf{H}^t$ . This is easy to show:

$$\begin{aligned} x_i \left( \sum_j 2H_{ij} x_j - 1 \right) &= x_i \sum_j (2H_{ij} x_j - x_j) \\ &= x_i \sum_j (H_{ij} x_j + (1 - H_{ji}) x_j - x_j) \\ &= x_i \sum_j (H_{ij} - H_{ji}) x_j \\ &= x_i \sum_j P_{ij} x_j \end{aligned}$$

If  $x^*$  is an equilibrium such that  $x_i^* > 0$  for all  $i$ , then  $\mathbf{P}x^* = 0$ , meaning that  $x^*$  is an eigenvector of  $\mathbf{P}$ , corresponding to a zero eigenvalue.

The formulation in terms of  $\mathbf{P}$  is also useful to connect to important results in game theory:  $\mathbf{P}$  is the payoff matrix for a two-player, zero-sum symmetric matrix game. Then, when  $\mathbf{H}$  is random, the game admits only one optimal strategy<sup>1,2</sup>, which in general is a mixed strategy composed of  $k$  pure strategies, with  $k$  odd. The optimal strategy corresponds to the equilibrium  $x^*$ . This means that, starting the dynamics with  $n$  species and any positive initial conditions, the same  $n - k$  species will go extinct, and the same  $k$  species will coexist.

Though we believe that mild deviations from zero-sum (for example, cases in which the number of individuals is fixed on average) would not change qualitatively our results, we maintain this assumption out of mathematical convenience, and because it allows us to connect directly with the afore-mentioned results in game theory.

#### S1.1.1 Neutral cycling

Next, we want to show that the species that coexist cycle neutrally around the single equilibrium point. To do so, we construct a Lyapunov function for the system and show that we can find a constant of motion for system—meaning that trajectories will follow closed orbits.

Suppose  $x_i^* > 0$  is the equilibrium of Eq. 1 of the main text. We write the function

$$V(x) = - \sum_i x_i^* \log \frac{x_i}{x_i^*}. \quad (\text{S7})$$

Because of Gibbs' inequality,  $V(x) \geq 0$  for any  $0 < x_i < 1$  and it is equal to zero only if  $x_i = x_i^*$  for all  $i$ . Note also that at equilibrium  $2 \sum_j H_{ij} x_j^* = 1$ . We write

$$\begin{aligned}
 \frac{dV}{dt} &= \sum_i \frac{\partial V}{\partial x_i} \frac{dx_i}{dt} \\
 &= - \sum_i \frac{x_i^*}{x_i} \frac{dx_i}{dt} \\
 &= -2 \sum_{i,j} x_i^* H_{ij} x_j + \sum_i x_i^* \\
 &= -2 \sum_{i,j} x_i^* H_{ij} x_j + 1 \\
 &= \sum_j \left( -2 \sum_i H_{ij} x_i^* \right) x_j + 1 \\
 &= \sum_j \left( -2 \sum_i (1 - H_{ji}) x_i^* \right) x_j + 1 \\
 &= \sum_j \left( -2 \sum_i x_i^* + 2 \sum_i H_{ji} x_i^* \right) x_j + 1 \\
 &= \sum_j (-2 + 1) x_j + 1 \\
 &= - \sum_j x_j + 1 \\
 &= 0.
 \end{aligned}$$

Thus, we have found a constant of motion, meaning that the system will follow closed orbits. Hence, unless we start the system precisely at  $x^*$ , the abundances will cycle neutrally around the equilibrium.

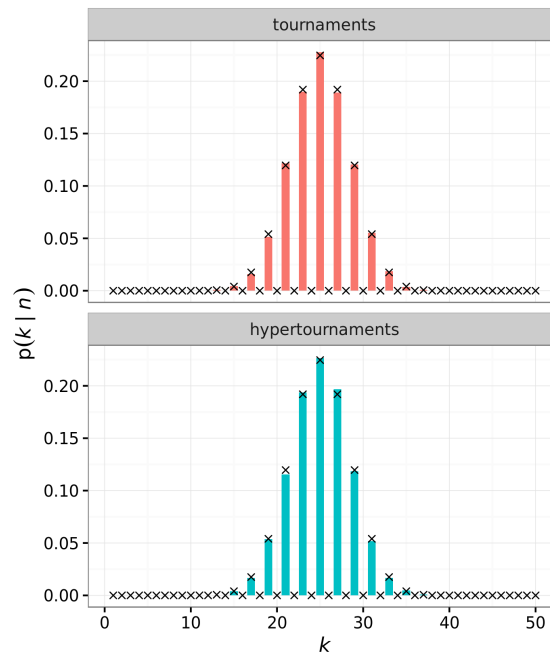
### S1.1.2 Number of coexisting species for random interactions

Now we ask how many species will coexist when we draw the matrix  $\mathbf{H}$  at random. We build random matrices in two ways. First, we set each pair  $(H_{ij}, H_{ji})$  to  $(1, 0)$  with probability  $1/2$ , and to  $(0, 1)$  with probability  $1/2$ ; the diagonal elements are all set to  $1/2$ . In this model, dominance is complete, and the matrix  $\mathbf{H}$  describes a tournament—a complete directed graph in which for each pair of species, an arrow connects the winner to the loser. Second, for each pair  $(H_{ij}, H_{ji})$  ( $i \neq j$ ) we sample a random number  $z$  from the uniform distribution  $\mathcal{U}[0, 1]$ , and set the values to  $(1 - z, z)$ ; again,  $H_{ii} = 1/2$  for all  $i$ . Thus  $\mathbf{H}$  encodes a generalization of tournament graphs, called hypertournaments<sup>3</sup>.

We set  $n = 50$ , and build 10,000 matrices of each kind. The number of coexisting species  $k$  can be found solving the corresponding linear program, as done by Allesina & Levine<sup>4</sup>. Finally, we tally the number of coexisting species to estimate  $p(k|n)$ . The results, reported in Fig. S2 show that both settings result in the same histogram:  $p(k|n) = 0$  when  $k$  is even, and  $p(k|n) = \binom{n}{k} 2^{1-n}$  when  $k$  is odd, as predicted both for random tournaments<sup>4,5</sup> and hypertournaments<sup>2</sup>.

### S1.1.3 Building $\mathbf{H}$ given $x^*$

Finally, we provide an elementary argument for why only an odd number of species coexist at equilibrium, unless we fine-tune parameters. The argument allows us to build an algorithm that



**Figure S2: Number of coexisting species with random competition.** Number of coexisting species  $k$  in a random tournament (red) or hypertournament (blue) when the number of initial species is  $n = 50$ . Bars: proportion of random (hyper)tournaments leading to the coexisting of  $k$  species (out of 10,000 simulations). Crosses: analytical expectation ( $p(k|n) = 0$  when  $k$  is even,  $\binom{n}{k} 2^{1-n}$  when  $k$  is odd).

accepts as input a desired species-abundance distribution  $x^*$ , and produces infinitely many matrices  $\mathbf{H}$  such that  $x^*$  is an equilibrium of Eq. 1, or Eq. 2 of the main text.

As we stated above, if all  $x_i^* > 0$  is an equilibrium, then  $x^*$  is an eigenvector of  $\mathbf{P}$  corresponding to a zero eigenvalue.

A matrix with one or more zero eigenvalues is rank-deficient. Fisher & Ryan<sup>6</sup> proved that, if the  $n \times n$  matrix  $\mathbf{H}$  represents a tournament, then  $\mathbf{P}$  has rank  $n$  when  $n$  is even, and rank  $n - 1$  when  $n$  is odd. Therefore, the matrix  $\mathbf{P}$  will have an eigenvalue of zero only if  $n$  is odd—when the matrix  $\mathbf{H}$  is a tournament matrix, only an odd number of species can coexist.

When we build the matrix  $\mathbf{H}$  at random by sampling the coefficients from a uniform distribution, we have that the matrix has full rank (i.e., the set of rank-deficient matrices has measure zero). For these matrices then, the same result found for tournaments holds<sup>2</sup>.

Clearly, one could build rank-deficient matrices—only, it is impossible to find this result at random when sampling the coefficients from a continuous distribution. In the next paragraphs, we show having a rank-deficient  $\mathbf{H}$  automatically introduces a neutral manifold of infinitely many equilibria, and that this case is unrobust: small perturbations of the coefficients will lead to an odd number of species coexisting.

Armed with these results, we can build an algorithm that takes a desired relative species abundance  $x^*$ , and builds a matrix  $\mathbf{H}$  such that Eq. 1 and Eq. 2 have  $x^*$  as an equilibrium. In fact, we show that one can construct infinitely many matrices of this kind. The algorithm also makes apparent the implicit assumptions we are making when we want an even number of species to coexist indefinitely.

The strategy is to build a matrix of eigenvectors  $\mathbf{U}$  for the skew-symmetric matrix  $\mathbf{P}$ , and the corresponding diagonal matrix of eigenvalues  $\mathbf{\Lambda}$ . Then,  $\mathbf{P} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ . Finally, we have that  $H_{ij} = (P_{ij} + 1)/2$ .

The first case we examine is that of an odd number of species,  $n$ . We set the real vector  $r_1 = x^*$ , and draw  $n - 1$  random, real-valued vectors,  $r_2, \dots, r_n$ . We modify the vectors  $r_2, \dots, r_n$  to make them orthogonal to  $r_1$  and each other. Writing  $U_j$  as the  $j^{\text{th}}$  column of  $\mathbf{U}$ , we set:  $U_1 = r_1 = x^*$ ,  $U_2 = r_2 + ir_3$ ,  $U_3 = r_2 - ir_3$ ,  $U_4 = r_4 + ir_5$ ,  $U_5 = r_4 - ir_5$ , etc. (note that  $i = \sqrt{-1}$ ). Finally, we need to set the eigenvalues. We choose  $\Lambda_{1,1} = 0$ , and we set the remaining eigenvalues in pairs: we draw a random number  $z$ , and set  $\Lambda_{2,2} = iz$ ,  $\Lambda_{3,3} = -iz$ ; draw a second random number to determine  $\Lambda_{4,4}$  and  $\Lambda_{5,5}$ , etc. Once we have chosen the eigenvalues and eigenvectors, we compute  $\mathbf{P}$ , and, if needed, normalize the matrix such that  $\min P_{ij} \geq -1$  and  $\max P_{ij} \leq 1$  (dividing all elements by a constant simply re-scales the spectrum).

Note that we have complete freedom in choosing all the eigenvalues and eigenvectors, besides the first. Hence, we can produce infinitely many matrices  $\mathbf{P}$  that have  $x^*$  as an eigenvector associated with a zero eigenvalue, and therefore we can produce infinitely many matrices  $\mathbf{H}$ .

If  $x^*$  contains an even number of species, we proceed in almost the same way, but we now need two eigenvalues to be 0 (because the matrix  $\mathbf{P}$  needs to be skew-symmetric, all the eigenvalues must have real part 0, and conjugate imaginary parts), and two eigenvectors to be real. Therefore, we construct the matrix  $\mathbf{U}$  as:  $U_1 = r_1 = x^*$ ,  $U_2 = r_2$ ,  $U_3 = r_3 + ir_4$ ,  $U_4 = r_3 - ir_4$ ,  $U_5 = r_5 + ir_6$ ,  $U_6 = r_5 - ir_6$ , etc. We set  $\Lambda_{1,1} = \Lambda_{2,2} = 0$ , and the remaining eigenvalues at random, as done for  $n$  odd. Again, we can build infinitely many  $\mathbf{P}$  and therefore  $\mathbf{H}$ .

This means that in order to have an equilibrium containing an even number of species, we need to set two eigenvalues to zero: the matrix has rank  $n - 2$ , so that the rows/columns of  $\mathbf{P}$  are no longer linearly independent. This is a very fragile situation, as small modifications of the coefficients of  $\mathbf{P}$  (or, equivalently,  $\mathbf{H}$ ) would break this fine-tuning. Hence, the dynamics around the equilibrium  $x^*$  are not robust. A more subtle point is that, given  $\mathbf{P}U_1 = \mathbf{P}U_2 = 0$ , any linear combination  $\mathbf{P}(\alpha U_1 + \beta U_2) = 0$  as well. Therefore, in the case of an even number of species coexisting, we have infinitely many equilibrium points, again stressing that this is a special (and biologically quite unrealistic) situation. Small changes to the matrix  $\mathbf{H}$  would invariably lead the system to collapse to neutral cycling around an equilibrium containing an odd number of species.

Code implementing this algorithm can be found at [git.io/vXZWF](https://github.com/vXZWF).

## S1.2 Sampling more than two seedlings

Under the hypothesis  $d_i = f_i = 1$ , Eq. S2 reduces to

$$\begin{cases} q_i^{(1)} = x_i \\ \dots \\ q_i^{(m)} = x_i \sum_k H_{ik} q_k^{(m-1)} + q_i^{(m-1)} \sum_k H_{ik} x_k \end{cases} \quad (\text{S8})$$

and Eq. S1 becomes

$$\frac{dx_i}{dt} = q_i^{(m)} - x_i. \quad (\text{S9})$$



We want to show that if  $x^*$  is the equilibrium of the model with two seedlings ( $m = 2$ ), it is also an equilibrium of the other models with  $m > 2$ . Consider Eq. S8 evaluated at  $x = x^*$ :

$$\begin{cases} q_i^{(1)*} = x_i^* \\ q_i^{(2)*} = 2x_i^* \sum_k H_{ik} x_k^* \\ \dots \\ q_i^{(m)*} = x_i^* \sum_k H_{ik} q_k^{(m-1)*} + q_i^{(m-1)*} \sum_k H_{ik} x_k^* \end{cases}$$

In Eq. S9 for  $m = 2$ , we have that  $\mathbf{H}x^* = 1/2$ . It is easily seen that if  $q_k^{(m-1)*} = x^*$ , then  $q_k^{(m)*} = x^*$ , and therefore, by induction, it follows that  $x^*$  is a fixed point of Eq. S9 for any  $m$ .

### S1.2.1 Sampling three seedlings

Now consider the system in Eq. S4, with identical physiological rate  $d_i = f_i = 1$ , describing the case of  $m = 3$

$$\frac{dx_i}{dt} = x_i \left( \sum_{j,k} (2H_{ij}H_{ik} + H_{ij}H_{jk} + H_{ik}H_{kj}) x_j x_k - 1 \right) = x_i \left( 2 \sum_{j,k} (H_{ij}H_{ik} + H_{ij}H_{jk}) x_j x_k - 1 \right), \quad (\text{S10})$$

which is Eq. 3 of the main text. This system of equations is equivalent to the replicator equation for a three-player game in Eq. 4 of the main text,

$$\frac{dx_i}{dt} = x_i \sum_{j,k} P_{ijk} x_j x_k, \quad (\text{S11})$$

where the coefficient  $P_{ijk} = 2H_{ij}H_{ik} - H_{ji}H_{jk} - H_{ki}H_{kj}$ . We write:

$$\begin{aligned} x_i \left( \sum_{j,k} (2H_{ij}H_{ik} + H_{ij}H_{jk} + H_{ik}H_{kj}) x_j x_k - 1 \right) &= x_i \sum_{j,k} (2H_{ij}H_{ik} + H_{ij}H_{jk} + H_{ik}H_{kj} - 1) x_j x_k \\ &= x_i \sum_{j,k} (2H_{ij}H_{ik} + (1 - H_{ji})H_{jk} + (1 - H_{ki})H_{kj} - 1) x_j x_k \\ &= x_i \sum_{j,k} (2H_{ij}H_{ik} - H_{ji}H_{jk} - H_{ki}H_{kj}) x_j x_k \\ &= x_i \sum_{j,k} P_{ijk} x_j x_k. \end{aligned}$$

Now we want to show that now an equilibrium  $x^* > 0$  is globally stable. Taking the same function  $V(x)$  defined in Eq. S7, we find

$$\begin{aligned}
 \frac{dV}{dt} &= -\sum_i \frac{x_i^*}{x_i} \frac{dx_i}{dt} \\
 &= -2 \sum_{i,j,k} (x_i^* H_{ij} H_{jk} x_j x_k + x_i^* H_{ij} H_{ik} x_j x_k) + \sum_i x_i^* \\
 &= -2 \sum_{j,k} \left( \sum_i x_i^* H_{ij} \right) H_{jk} x_j x_k - 2 \sum_i x_i^* \left( \sum_j H_{ij} x_j \right)^2 + 1 \\
 &= -2 \sum_{j,k} \frac{1}{2} H_{jk} x_j x_k - 2 \sum_i x_i^* \left( \sum_j H_{ij} x_j \right)^2 + 1 \\
 &= -\frac{1}{2} - 2 \sum_i x_i^* \left( \sum_j H_{ij} x_j \right)^2 + 1 \\
 &= -2 \sum_i x_i^* \left( \sum_j H_{ij} x_j \right)^2 + \frac{1}{2},
 \end{aligned}$$

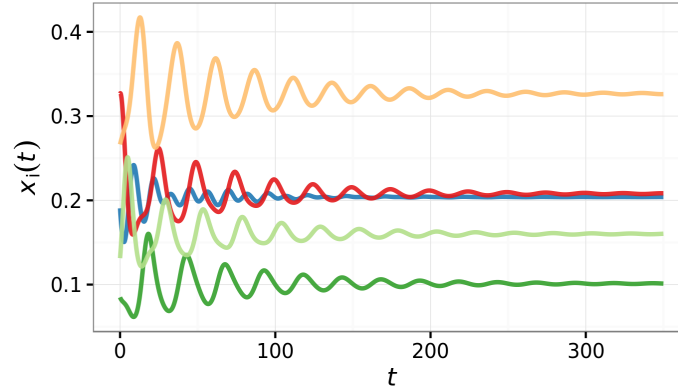
where we used  $\sum_i x_i^* H_{ij} = 1/2$  and  $\sum_{j,k} H_{jk} x_j x_k = 1/2$ . Next we introduce  $\xi_j := x_j - x_j^*$  (note that  $\sum_j \xi_j = 0$  by definition), obtaining

$$\begin{aligned}
 \frac{dV}{dt} &= -2 \sum_i x_i^* \left( \sum_j H_{ij} (x_j^* + \xi_j) \right)^2 + \frac{1}{2} \\
 &= -2 \sum_i x_i^* \left( \frac{1}{2} + \sum_j H_{ij} \xi_j \right)^2 + \frac{1}{2} \\
 &= -\frac{1}{2} - 2 \sum_i x_i^* \sum_j H_{ij} \xi_j - 2 \sum_i x_i^* \left( \sum_j H_{ij} \xi_j \right)^2 + \frac{1}{2} \\
 &= -2 \sum_j \left( \sum_i x_i^* H_{ij} \right) \xi_j - 2 \sum_i x_i^* \left( \sum_j H_{ij} \xi_j \right)^2 \\
 &= -\sum_j \xi_j - 2 \sum_i x_i^* \left( \sum_j H_{ij} \xi_j \right)^2 \\
 &= -2 \sum_i x_i^* \left( \sum_j H_{ij} \xi_j \right)^2 \leq 0
 \end{aligned}$$

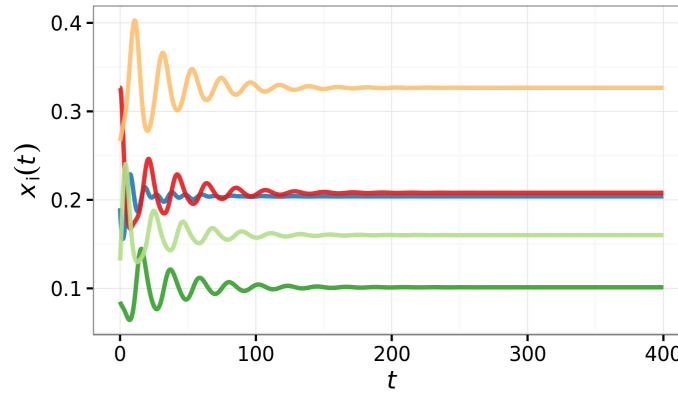
for any choice of  $\xi$ , and therefore, for any value of  $x$ . Assuming that the matrix  $\mathbf{H}$  is of full rank,  $dV/dt = 0$  only if  $\xi = 0$ , i.e., only if  $x = x^*$ . Since  $V(x) \geq 0$  for any  $x$  and  $V(x) = 0$  only if  $x = x^*$ ,  $dV/dt \leq 0$  implies that  $x = x^*$  is a globally stable fixed point. Therefore, in the model where we sample three seedlings we have global convergence—starting the system at any initial condition leads to the same outcome, unless we have a rank-deficient  $\mathbf{H}$  (e.g., the case in which we want an even number of species to coexist, in which case the system will reach one of the infinitely many equilibria).

When we take more than three seedlings at a time, the results are qualitatively the same, but convergence to the equilibrium is faster (Fig. S3).

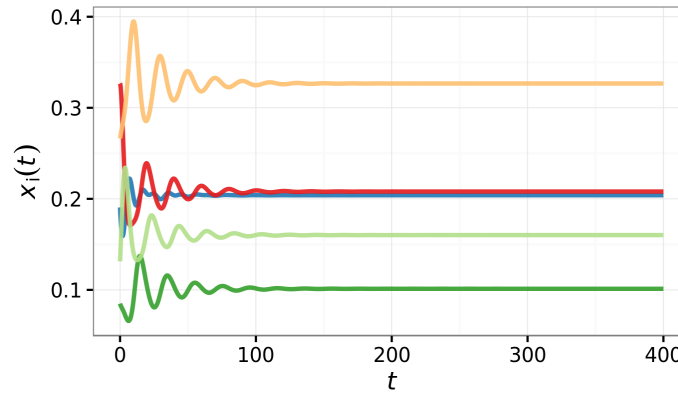
**a) Sampling three seedlings**



**b) Sampling four seedlings**



**c) Sampling five seedlings**



**Figure S3: Sampling more than three seedlings accelerates the convergence to the equilibrium.** We took the system in Fig. 2d of the main text, and integrated the dynamics when we compete three (a), four (b), or five (c) seedlings at a time. We plot the relative abundance of each species (y-axis) at time  $t$  (x-axis).

## S2 Different physiological rates

### S2.1 Sampling two seedlings

In this section, we explicitly study Eq. S3, which can be rewritten as

$$\frac{dx_i}{dt} = \frac{x_i}{F(x)^2} \left( D(x) \left( \sum_k f_i 2H_{ik} f_k x_k \right) - d_i F(x)^2 \right). \quad (\text{S12})$$

If a feasible solution  $x^* > 0$  exists, it is the solution of

$$\sum_k H_{ik} f_k x_k^* = \frac{d_i F(x^*)^2}{f_i 2D(x^*)}. \quad (\text{S13})$$

The equation has the form:

$$\sum_k H_{ik} f_k x_k^* = c \frac{d_i}{f_i}, \quad (\text{S14})$$

whose solution is

$$x_i^* = c \frac{1}{f_i} \sum_k H_{ik}^{-1} \frac{d_k}{f_k}. \quad (\text{S15})$$

The solution must satisfy  $\sum_i x_i^* = 1$ , and therefore we can set the normalization constant  $c$ :

$$x_i^* = \frac{1}{f_i \sum_{jk} H_{jk}^{-1} \frac{d_k}{f_j f_k}} \sum_k H_{ik}^{-1} \frac{d_k}{f_k}. \quad (\text{S16})$$

Having characterized the equilibrium, we turn to its stability. Assuming that all  $x_i^* > 0$  (i.e., that the fixed point is feasible), we introduce the function  $V(x)$  defined in Eq. S7, obtaining

$$\begin{aligned} \frac{dV}{dt} &= \sum_i \frac{\partial V}{\partial x_i} \frac{dx_i}{dt} \\ &= - \sum_i \frac{x_i^*}{x_i} \frac{dx_i}{dt} \\ &= - \frac{1}{F(x)^2} \left( D(x) \left( 2 \sum_{ik} x_i^* f_i H_{ik} f_k x_k \right) - F(x)^2 \sum_i x_i^* d_i \right). \end{aligned} \quad (\text{S17})$$

Introducing  $H_{ik} = 1 - H_{ki}$  in Eq. S13, we find

$$2 \sum_i x_i^* f_i H_{ik} = 2F(x^*) - \frac{d_k F(x^*)^2}{f_k D(x^*)}, \quad (\text{S18})$$

and therefore

$$\begin{aligned} \frac{dV}{dt} &= - \frac{1}{F(x)^2} \left( 2D(x)F(x^*)F(x) - D(x)^2 \frac{F(x^*)^2}{D(x^*)} - F(x)^2 D(x^*) \right) \\ &= - \frac{1}{F(x)^2 D(x^*)} (2D(x)F(x^*)F(x)D(x^*) - D(x)^2 F(x^*)^2 - F(x)^2 D(x^*)^2) \\ &= \frac{1}{F(x)^2 D(x^*)} (D(x)F(x^*) - F(x)D(x^*))^2 \geq 0 \end{aligned} \quad (\text{S19})$$

for any choice of  $x$ . This implies that any feasible fixed point (i.e., any positive solution of Eq. S13) is either neutrally stable or unstable. Neutral stability (i.e.,  $dV/dt = 0$ ) is achievable if and only if the ratio  $d_i/f_i$  has the same value for all the  $i$ . Consistent with this finding, Fig. S4 shows that whenever physiological parameters are perturbed, the cycles become unstable, eventually leading to monodominance.

## S2.2 Sampling three seedlings

Finally, we study Eq. S4, that can be rewritten as

$$\frac{dx_i}{dt} = \frac{x_i}{F(x)^3} \left( D(x) \sum_{jk} f_i 2H_{ij} (H_{ik} + H_{jk}) f_j x_j f_k x_k - d_i F(x)^3 \right). \quad (\text{S20})$$

If a feasible solution  $x^* > 0$  exists, it is the solution of

$$\sum_{jk} H_{ij} (H_{ik} + H_{jk}) f_j x_j^* f_k x_k^* = \frac{d_i F(x^*)^3}{f_i 2D(x^*)}. \quad (\text{S21})$$

Notice that the  $x^*$  for the case of two seedlings does not solve this equation—the equilibrium depends on the number of seedlings sampled, in contrast to what found when we assumed all physiological rates to be the same.

Assuming that a feasible solution of Eq. S21 exists, we write the Lyapunov function  $V(x)$  (Eq. S7):

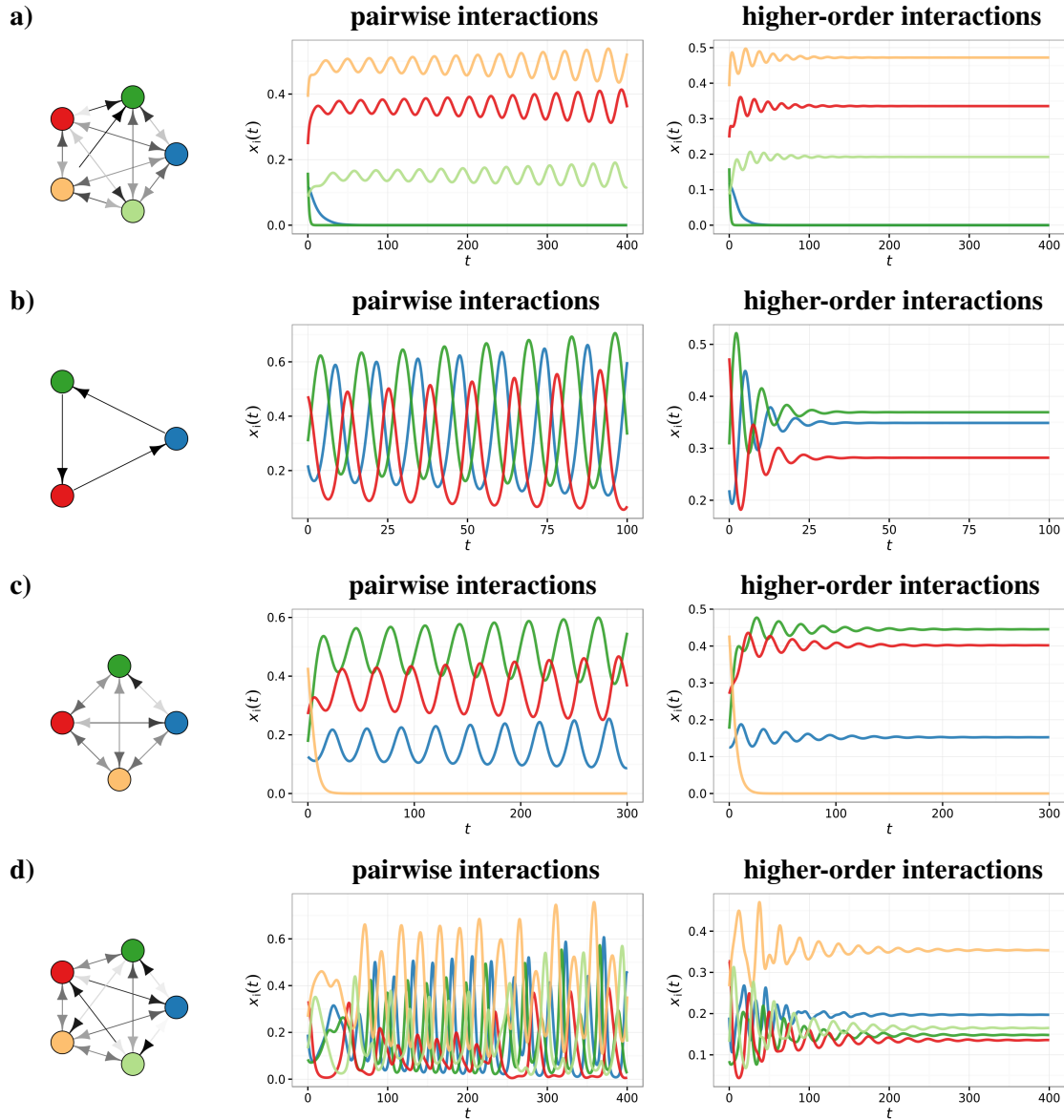
$$\begin{aligned} \frac{dV}{dt} &= \sum_i \frac{\partial V}{\partial x_i} \frac{dx_i}{dt} = - \sum_i \frac{x_i^*}{x_i} \frac{dx_i}{dt} \\ &= - \frac{1}{F(x)^3} \left( D(x) \left( 2 \sum_{ijk} x_i^* f_i H_{ij} (H_{ik} + H_{jk}) f_j x_j f_k x_k \right) - F(x)^3 \sum_i x_i^* d_i \right). \end{aligned} \quad (\text{S22})$$

After some lengthy calculation, we find the expression

$$\begin{aligned} \frac{dV}{dt} &= \frac{1}{F(x)^3 D(x^*)} (F(x^*)^3 D(x)^2 - 2F(x^*)^2 F(x) D(x) D(x^*) + F(x)^3 D(x^*)^2) + \\ &\quad - \frac{2}{F(x)^3} \sum_{ijk} x_i^* f_i H_{ij} f_j \xi_j H_{jk} f_k \xi_k - \frac{2}{F(x)^3} \sum_{ik} x_i^* f_i (H_{ik} f_j \xi_j)^2 \end{aligned} \quad (\text{S23})$$

that can be interpreted more easily. Take the case of  $f_i = d_i = 1$ : then, only the last term is different from zero, and since it is always non-positive, we find that the fixed point is globally stable. The two terms that appear when physiological rates are different do not have a definite sign. This implies that, at least in principle, for some values of parameters the fixed point could be non-attractive. However, note two facts: 1) dynamics are stable for  $f_i = d_i = 1$ , and 2) the position of the fixed point and the derivative of the Lyapunov function are both continuous functions of the physiological rates. It follows that for a sufficiently small perturbation of the  $f_i$  and the  $d_i$  away from 1 will not be able to make a qualitative change to the dynamical properties of the model. Numerical simulations (Fig. S4, and next section) show that this is indeed the case.

In summary, whether the physiological rates are allowed to vary or not, the case in which we sample three seedlings is very different from that of sampling two seedlings. When we sample only two seedlings, coexistence is always transient, while when we sample three seedlings, we can have stable coexistence.



**Figure S4: Dynamics with variable physiological rates.** When introducing small changes to the physiological rates ( $d_i$  and  $f_i$ ), the equilibrium become unstable for pairwise interactions: the system cycles away from equilibrium, eventually leading to monodominance. For the same system in Fig. 1 of the main text, we introduce different  $d_i$  and  $f_i$  sampling them randomly from  $\mathcal{U}[0.9, 1.1]$ . Note that in all cases for pairwise interactions the amplitude of the cycles increases with time, and that the dynamically-fragile case of the coexistence of an even-number of species (c) is immediately broken. For higher order interactions, the dynamics are robust to small changes in the parameterization. In this case, we recover the same qualitative dynamics found for the case of all identical physiological rates (Fig. 2 of the main text), besides the fact that the coexistence between an even number of species is broken.

### S2.2.1 Numerical analysis

We expect the dynamics of Eq. S3 to be very rich, including parameter and initial conditions combinations leading to fixed points, limit cycles, and possibly chaos. Though a full analytic characterization of this system of equations will be difficult to achieve, from an ecological point of view two questions are important: a) can many species coexist when we allow physiological rates to vary?, and b) can we reach coexistence from a variety of initial conditions? (i.e., is the basin of attraction of coexistence large enough?). We use numerical simulations to attempt answering these questions.

The two numerical experiments follow the same basic setting. 1) We sample the relevant parameters: for a given number of initial species,  $n$ , the matrix  $\mathbf{H}$  is drawn at random, the fertility values  $f_i$  are independently sampled from the uniform distribution  $\mathcal{U}[1 - \nu, 1 + \nu]$ , where the parameter  $\nu$  modulates the variability in the rates;  $d_i$  is obtained multiplying  $f_i z_i$ , where  $z_i$  is sampled from the same distribution (hence,  $d_i/f_i$  is uniformly distributed, while  $d_i$  is not). Finally, initial conditions  $x_i(0) > 0$  are set at random. 2) The dynamics are integrated numerically for a long time ( $5 \cdot 10^6$  time units), so that in most cases the transients should have elapsed. 3) We record the number of species coexisting at the end of the simulation ( $k$ ), and their identity.

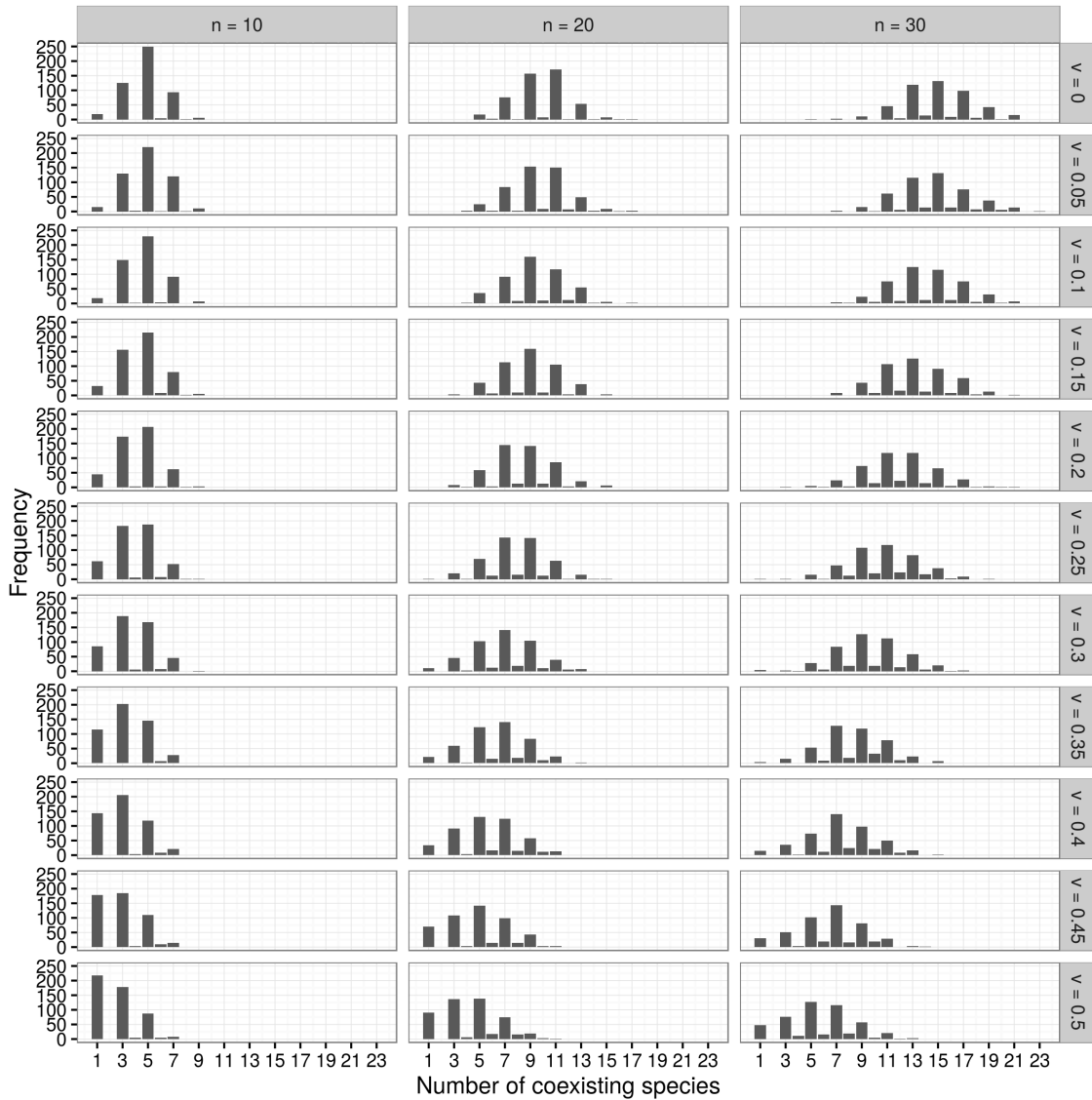
There are two main limitations of this approach: first, because of the inevitable rounding errors, we could declare a species extinct when it approaches zero density (though mathematically it could eventually rebound); second, the opposite problem could also be present for certain parameterizations: given the arbitrary length of integration, we could have species that are slowly but steadily going towards extinction be declared extant in our calculations. We expect the effect of both problems to be relatively small, such that our numerical simulations should approximate the actual dynamics quite closely. This expectation is confirmed when inspecting the numerical results for  $\nu = 0$ , in which case we should recover our analytical results.

In our first experiment, we take  $n \in \{10, 20, 30\}$  and  $\nu \in \{0, 0.05, 0.1, \dots, 0.5\}$  and produce 500 simulations for each combination of parameters. The histograms showing the frequency of the number of coexisting species  $k$  for each parameter combinations are reported in Fig. S5. For  $\nu = 0$  (and hence identical physiological rates), the histograms are very similar to the analytic results in Fig. S2, though in a few cases we still have an even number of species, meaning that for some of the simulations the transients are very long. When we increase  $\nu$ , the average number of coexisting species is reduced, but the decrease is slow and fairly linear, meaning that even when physiological rates are quite different (e.g.,  $\nu = 0.5$ ), we still have a fairly high proportion of species coexisting.

In the second numerical experiment we probe the size of the basin of attraction. As before, we vary  $n$  and  $\nu$  to test the effect of the variability on physiological rates. Differently from the other experiment, for each combination of  $n$  and  $\nu$ , we choose 100 parameterizations (i.e., setting  $\mathbf{H}$ ,  $d$  and  $f$ ), and then we integrate the dynamics 100 times, starting from different initial conditions. We then record how often the same set of species coexist at the end of the dynamics. A proportion of 1 means that for all 100 initial conditions, we always found that the same set of species coexist at the end of the simulation. A proportion lower than 1 means that depending on the initial conditions, we end up with alternative outcomes. Out of the 1000 parameterizations (Fig S6), in 848 cases we found that the endpoint was exactly the same for all 100 initial conditions (corresponding to a proportion of 1). In only 75 cases out of 1000, we found a proportion  $< 0.9$  (meaning that at fewer than 90 initial conditions led to the same outcome).

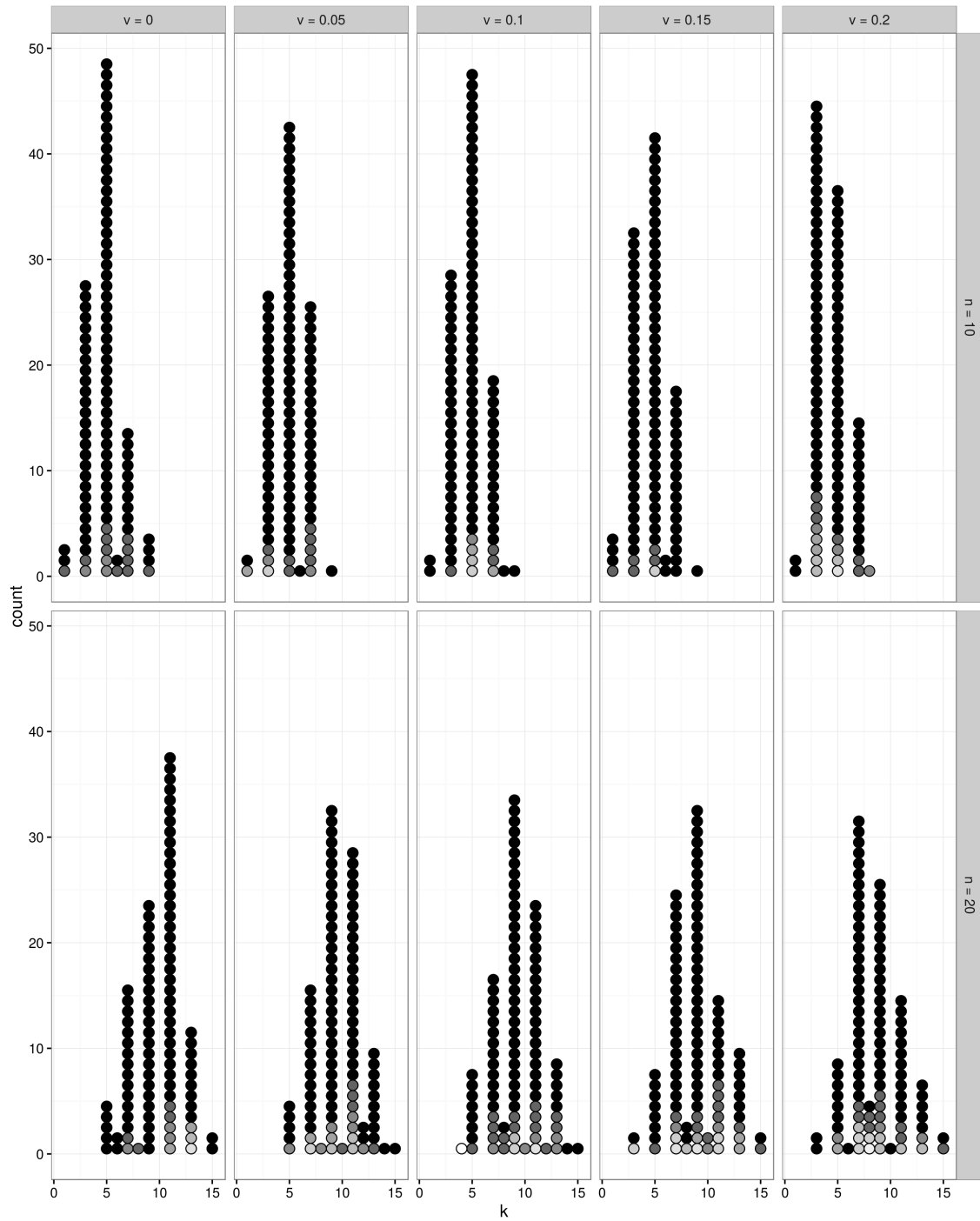
In summary, these numerical results confirm the intuition we built analytically: even when considering different physiological rates, the model allows for the long-term coexistence of many

species, with the attractor having a large basin. To lower coexistence levels, naturally one can make physiological rates so different from each other that species are excluded, exactly as expected from ecological considerations.



**Figure S5: Number of coexisting competitors when physiological rates are different for all species.** We sampled independently both  $d_i$  and  $z_i = d_i/f_i$  from the uniform distribution  $\mathcal{U}[1-v, 1+v]$ , and varied  $v$  to explore how the variance in these rates affects coexistence (rows). For each choice of initial number of species ( $n$ , columns), we generated a random matrix  $\mathbf{H}$ , random initial conditions, and integrated the dynamics in Eq. S4 for a long time ( $5 \cdot 10^6$  steps). We recorded the number species coexisting at the end of the simulation, and produced a histogram by repeating the procedure 500 times for any choice of  $n$  and  $v$ . For  $v = 0$ , the results should closely match those in Fig. S2. The results show that the histogram gently shifts to the left when increasing the variability of the physiological rates  $v$ —as long as the physiological parameters are similar enough, we can have the coexistence of many species.





**Figure S6: Dependence of the attractor on initial conditions.** For  $n = 10$  or  $20$  (rows), and for different values of  $v$ , we repeat the simulations in Fig. S5 but starting from 100 random initial conditions. The position of the dots in the histogram corresponds to the number of species coexisting at the end of the simulations ( $k$ ), and the color to the probability of ending up with the same  $k$  species coexisting. Black corresponds to a probability of 1, and the lightest shade of gray ( $n = 20$ ,  $v = 0.1$ ,  $k = 4$ ) to a probability of 29%. In only 75 cases out of 1000 we found a probability  $< 0.9$ .

### S3 Intransitivity and coexistence

In the theory studied by Allesina & Levine<sup>4</sup>—where dominance is complete and the matrix  $\mathbf{H}$  encodes a tournament—species can coexist only if they are connected by an intransitive cycle such as in the rock-paper-scissors game. The existence of intransitive cycles in competitive abilities of species has found mixed evidence in empirical data, with few clear examples in marine sessile organisms<sup>7–9</sup>, and in microbial systems<sup>10,11</sup>. In terrestrial plants, the consensus seems to favor strict competitive hierarchies<sup>12,13</sup>, though recent statistical models suggest diffuse intransitive competition in speciose communities<sup>14</sup>.

Allesina & Levine<sup>4</sup> showed that intransitivity can arise whenever there are multiple resources, and species experience a trade-off such that they cannot excel at competing for all resources simultaneously. Similarly, intransitivity can be due to spatial/temporal heterogeneity (e.g., one species wins in the shade, the other in the open ground; one in the wet season, one in the dry), trophic interactions (herbivores consuming preferentially strong competitors), and many other processes.

When we relax the strict requirement of complete dominance, but keep the physiological parameters constant among species, we still require intransitivity—though now its definition is probabilistic. In particular, we conjecture that whenever  $k$  species coexist, they are connected through a cycle in the graph defined by the coarse-grained matrix  $\mathbf{H}'$  obtained by setting  $H'_{ij} = 1$  whenever  $H_{ij} > 1/2$  and  $H'_{ij} = 0$  otherwise (as always, we assume  $\mathbf{H}$  to be of full rank, and the coefficients to be sampled from a continuous distribution, to avoid dealing with the case of  $H_{ij}$  being exactly  $1/2$  for  $i \neq j$ ). Note however that the reverse is not true: the fact that the species are connected through a cycle in  $\mathbf{H}'$  does not guarantee coexistence, as shown by the following simple example. Take  $\mathbf{H}$  to be:

$$\mathbf{H} = \begin{bmatrix} 0.50 & \underline{0.63} & \underline{0.79} & 0.18 & 0.47 \\ 0.37 & 0.50 & \underline{0.58} & \underline{0.93} & 0.48 \\ 0.21 & 0.42 & 0.50 & \underline{0.99} & \underline{0.84} \\ \underline{0.82} & 0.07 & 0.01 & 0.50 & \underline{0.66} \\ \underline{0.53} & \underline{0.52} & 0.16 & 0.34 & 0.50 \end{bmatrix} \quad (\text{S24})$$

where we underlined all the coefficients  $> 1/2$ . Solving for the equilibrium in the case in which we sample two or more seedlings, we find  $x^* \approx \{0.49, 0.36, 0, 0.15, 0\}$ , meaning that species 3 and 5 will go extinct. Note that species 1, 2, and 4 are connected by the “intransitive” cycle  $H_{1,2}H_{2,4}H_{4,1} = 0.63 \cdot 0.93 \cdot 0.82$  with all coefficients  $> 1/2$ . When we coarse-grain the matrix we obtain the matrix  $\mathbf{H}'$ :

$$\mathbf{H}' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (\text{S25})$$

in which two cycles connect all the species, yielding a prediction of  $x'^*$  in which all the species would have positive density ( $x'_i = 1/5$  for all  $i$ )—examining  $\mathbf{H}'$  we would expect coexistence among all species, but analyzing the matrix  $\mathbf{H}$  we predict the extinction of two species.

In summary, for identical physiological rates intransitivity is defined only probabilistically, but still plays a role: all the species coexisting at equilibrium are connected through a cycle whose coefficients are all greater than  $1/2$ . The existence of such a cycle, however, does not guarantee coexistence.

Next, we show that when we allow physiological parameters to differ between species intransitivity is not a necessary condition for coexistence. That is, a species can persist despite losing competitive bouts with each and every other species more often than not. Given that we cannot solve analytically for the equilibrium in this more general case, we report here just a few numerical examples that however are sufficient to illustrate this surprising result. Take for example the matrix:

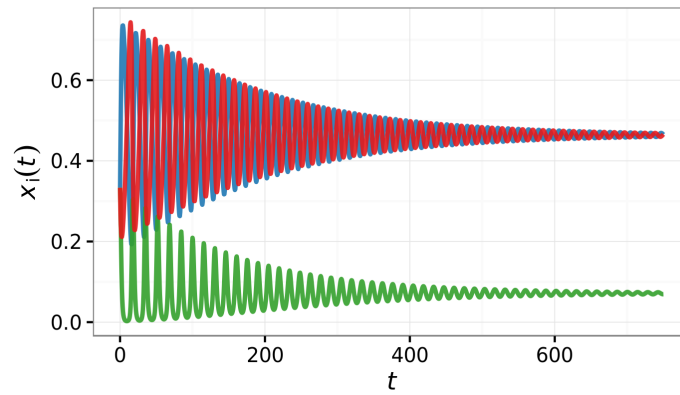
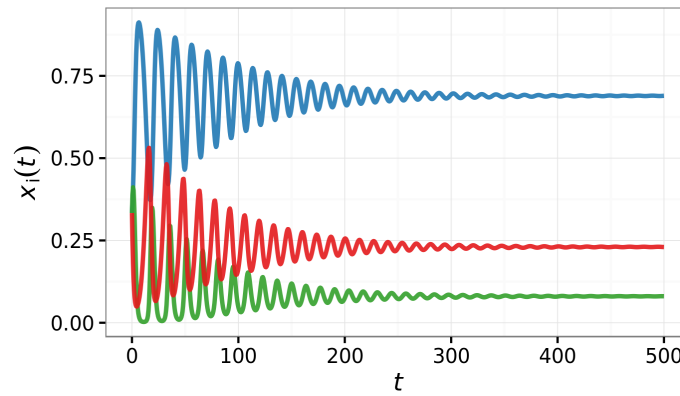
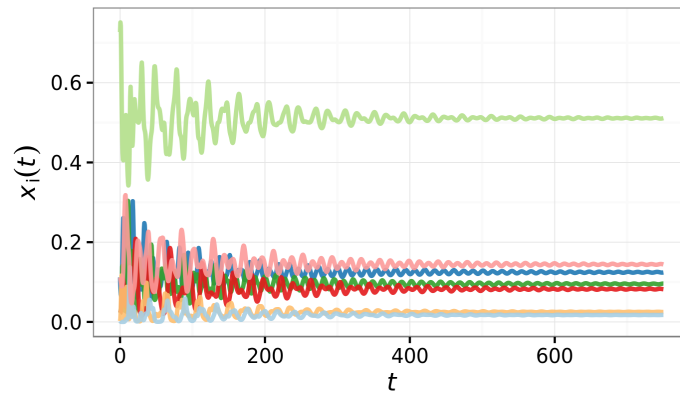
$$\mathbf{H} = \begin{bmatrix} 0.50 & \underline{1.00} & \underline{0.62} \\ 0.00 & 0.50 & \underline{1.00} \\ 0.38 & 0.00 & 0.50 \end{bmatrix} \quad (\text{S26})$$

which is clearly transitive in probability—we can gather all the coefficients with value  $> 1/2$  in the upper-triangular part of the matrix. Yet, all species persist at equilibrium when for example we set  $f = \{1, 1, 1\}$  and  $d = \{7/4, 5/4, 1\}$  (Fig. S7a), or  $f = \{1/2, 1, 1\}$  and  $d = \{1, 1, 1\}$  (Fig. S7b).

Searching the parameter space, we can find larger communities that can persist despite a matrix  $\mathbf{H}$  that is transitive in probability. For example, Fig. S7c shows the persistence at equilibrium of a seven-species community with:

$$\mathbf{H} = \begin{bmatrix} 0.50 & 0.76 & \underline{0.52} & \underline{0.68} & \underline{0.54} & \underline{0.51} & \underline{0.56} \\ 0.24 & 0.50 & \underline{0.59} & \underline{0.74} & \underline{0.71} & \underline{0.62} & \underline{0.77} \\ 0.48 & 0.41 & 0.50 & \underline{0.92} & \underline{0.74} & \underline{0.82} & \underline{0.56} \\ 0.32 & 0.26 & 0.08 & 0.50 & \underline{0.97} & \underline{0.76} & \underline{0.70} \\ 0.46 & 0.29 & 0.26 & 0.03 & 0.50 & \underline{1.00} & \underline{0.82} \\ 0.49 & 0.38 & 0.18 & 0.24 & 0.00 & 0.50 & \underline{0.88} \\ 0.44 & 0.23 & 0.44 & 0.30 & 0.18 & 0.12 & 0.50 \end{bmatrix} \quad (\text{S27})$$

and physiological parameters  $d = \{4.5, 5, 3, 6.5, 0.6, 6.0, 1.3\}$  and  $f = \{3.7, 4.5, 2.5, 7, 0.7, 9, 2.5\}$ . These examples prove that intransitivity (in probability) is not a necessary condition for coexistence when species have different physiological rates.

**a) Matrix  $H$  as in Eq. S26,  $f = \{1, 1, 1\}$  and  $d = \{7/4, 5/4, 1\}$** **b) Matrix  $H$  as in Eq. S26,  $f = \{1/2, 1, 1\}$  and  $d = \{1, 1, 1\}$** **c) Matrix  $H$  as in Eq. S27**

**Figure S7: Dynamics under hierarchical competition.** When we allow species to have different physiological parameters, intransitivity is not necessary for coexistence. These time series are obtained for systems in which we can order the species such that each species has a probability of winning  $H_{ij} > 1/2$  for each  $i > j$  (i.e., the matrix is transitive in probability). Panels a) and b) are obtained for the matrix in Eq. S26, while panel c) for the larger system in Eq. S27. We plot the relative abundance of each species (y-axis) at time  $t$  (x-axis).

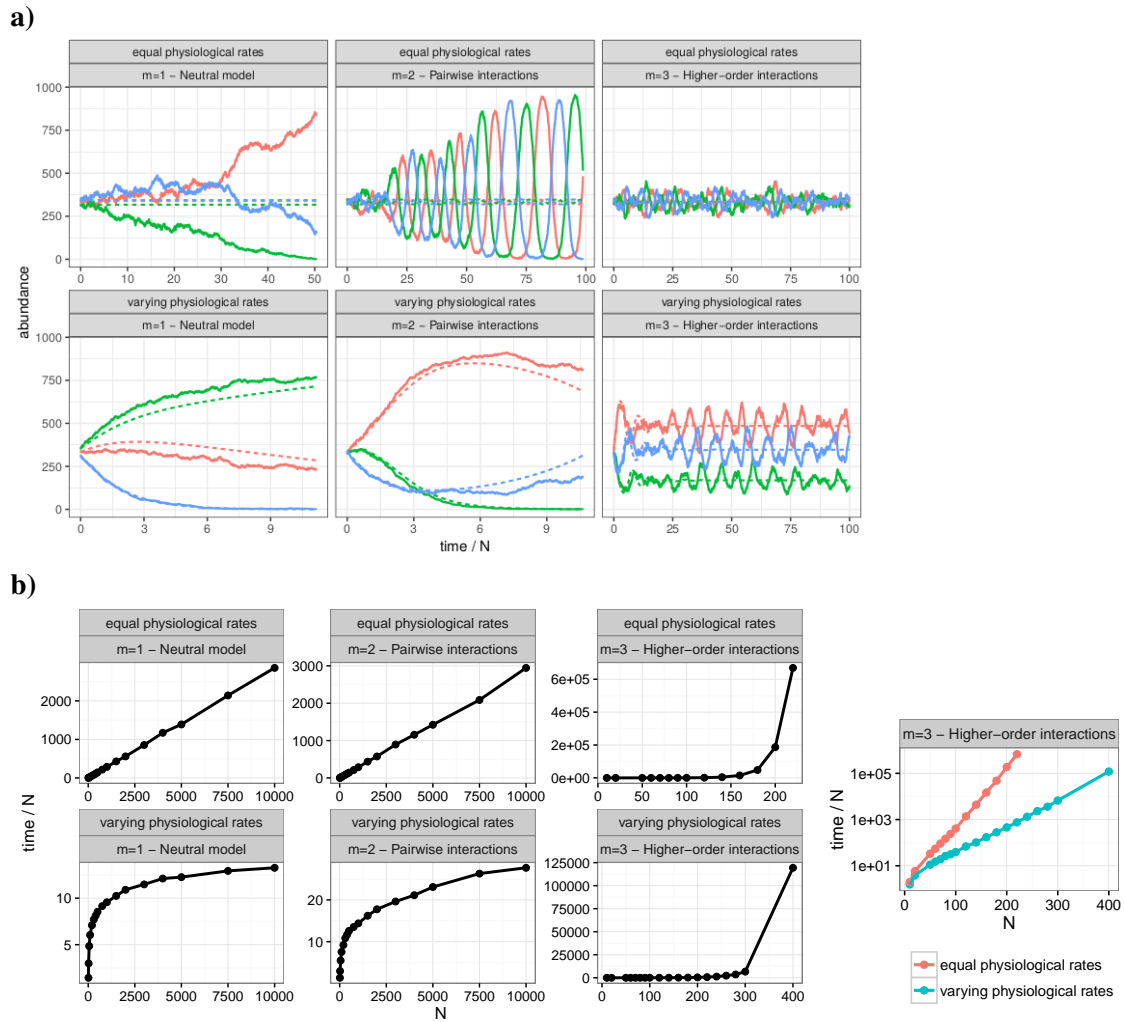
## S4 Stochastic models

So far, we have studied communities whose dynamics are well-described by systems of differential equations, assuming a large number of individuals ( $N \rightarrow \infty$ ). When the number of individuals is finite, demographic stochasticity plays an important role, sometimes producing non trivial effects. To test whether higher-order interactions influence dynamics in this setting, we explicitly simulate the stochastic dynamics defined by the rates  $W_{ij}^{(m)} = d_j x_j q_i^{(m)}$ .

Figure S8 shows the stochastic trajectories for a different number of competing individuals ( $m = 1, 2$ , or  $3$ ). The case  $m = 1$  corresponds to a neutral model (in absence of variability of physiological rates),  $m = 2$  to pairwise interactions, and  $m \geq 3$  to higher-order interactions. The deterministic analysis described in the previous sections is expected to be exact in the limit  $N \rightarrow \infty$ . In general, if noise is not additive, as in the case of demographic stochasticity, the full stochastic dynamical behavior cannot be predicted from the deterministic approximation. There are in fact many examples of models in which stochasticity changes the stability properties of the solutions. For example, Biancalani *et al.*<sup>15</sup>, have shown that noise not only governs the transition between alternative stable states, but can also generate them; closer to the theme of this work, Capitán *et al.*<sup>16</sup> have shown that the presence of noise makes competitive Lotka-Volterra system lose species at lower values of similarity between species compared to what expected under deterministic dynamics. In this section we show that, for finite  $N$ , the importance of stochastic fluctuations depends (mainly, but not exclusively) on the stability properties of the deterministic dynamics.

In the case of neutral stability, stochasticity produce non-trivial effects and the stochastic trajectories deviate from the deterministic predictions. In our model, neutral stability is found when physiological parameters are all identical, and  $m$  is either 1 (neutral case), or 2 (pairwise interactions). When on the other hand we have a strong deterministic driver, stochastic fluctuations simply produce fluctuations around the fixed point. For instance if higher-order interactions are considered ( $m \geq 3$ ), the presence of an attractive fixed point constrains trajectories to fluctuate around the equilibrium (Fig. S8). Finally, when the fixed point is unstable, deterministic dynamics drive species to extinction.

More quantitatively, the interplay between deterministic properties and stochasticity can be analyzed considering the scaling of extinction times with the population size. Fig. S8 shows time of extinction  $T_N$  (averaged over 1000 realizations with random initial conditions) for different choices of parameters and total population sizes  $N$ . When the fixed point is neutrally stable ( $m = 1$  and  $m = 2$ , with identical physiological rates)  $T_N/N$  is approximately linear in  $N$ . When the fixed point is unstable ( $m = 1$  and  $m = 2$  with different physiological rates),  $T_N/N$  becomes sub-linear in  $N$ , a scaling that reflects the deterministic dynamics driving one species to extinction. In the case of a stable fixed point ( $m = 3$ )  $T_N/N$  grows exponentially with  $N$  (see Fig. S8), implying the existence of a well defined (meta)-stable state.



**Figure S8: Stochastic simulations without immigration.** a) Stochastic trajectories for a rock-paper-scissors system, when we vary the number of seedlings competing. Dotted lines correspond to the deterministic approximation. For  $m = 1$  and equal physiological rates, one recovers a neutral model. For equal physiological rates and pairwise interactions ( $m = 2$ ) stability is only neutral and the deterministic solution does not describe the stochastic trajectories. Densities fluctuate around the fixed point, but the amplitude of these oscillations is driven by stochasticity, rather than being fixed as in the deterministic case. For higher-order interactions ( $m = 3$ ), densities fluctuate around the stable equilibrium of the deterministic system. When physiological rates are different, the stochastic trajectories are predicted by the deterministic equations. For  $m \leq 2$ , extinctions occur very rapidly, and are driven by the deterministic dynamics. For  $m = 3$ , trajectories fluctuate around the stable equilibrium. In all the simulations the total population abundance was set to  $N = 1000$ . b) Scaling of the average time to first extinction ( $T_N$ ) vs. total population abundance  $N$ . For equal physiological rates and  $m \leq 2$ , extinction time scales approximately linearly with  $N$ , as expected in neutral models as well as in the case of pairwise interactions<sup>17</sup>. When three or more seedlings are sampled, time to extinction grows exponentially with  $N$ . This is a consequence of the stability of the fixed point. When physiological rates are allowed to vary and  $m \leq 2$ ,  $T_N/N$  is very small, and scales logarithmically with  $N$ . The logarithmic scaling is a consequence of the instability of the deterministic equations. In agreement with the deterministic analysis, the stability of the fixed point in the case  $m \geq 3$  produces an exponential scaling of  $T_N/N$  (see also panel on the bottom right). In the case of varying physiological rates we considered  $f = (1., 1.2, 0.8)$  and  $d = (0.9, 1., 1.3)$ . Note that the scales of the  $x$  and  $y$  axis are different between different panels.

## S5 Stochastic models with immigration

Since its inception<sup>18,19</sup> neutral theory has been set in a stochastic framework in which the number of individuals is finite, and new species can be introduced by either speciation or immigration. Rich communities are then built by balancing the inevitable extinction processes with the introduction of new species.

Here we perform simulations that provide a direct comparison of our models with neutral models, allowing us to test whether we can retain some of their strengths while overcoming some of their limitations. Early critiques of neutral theory focused on the assumption of ecological equivalence and zero-sum dynamics<sup>20</sup>. However, theories should be judged by their predictions, rather than assumptions, and in time neutral theory proved to be strong in at least three aspects: a) it can produce species-rich communities<sup>21–23</sup>; b) it produces species-abundance distributions that resemble those observed empirically<sup>21,23–25</sup>; and c) it is mathematically tractable<sup>21–24</sup>. Recently, however several authors highlighted how neutral models fail in at least two aspects: first, species' abundances fluctuations seem to be too modest, with respect to what observed in natural populations<sup>26</sup>; second, because species are performing what amounts to a random walk in the space of abundances, very abundant species are more likely to be very old—in contrast to empirical evidence<sup>27–29</sup>.

In this section, we consider simulations in which the number of individuals  $N$  is finite—therefore explicitly considering demographic stochasticity—and new species can enter the system (through speciation or immigration from a metacommunity species pool) with probability  $\nu$ . In this context, a series of articles by McKane, Alonso and Solé<sup>22,30,31</sup> have investigated the dynamics of systems in which species interact in pairs, and dominance is either complete ( $S_{ij} > 0$  and  $S_{ji} < 0$  in the notation of their first contribution<sup>22</sup>), or neutral ( $S_{ij} = S_{ji} = 0$ ). Our approach to simulating community assembly closely matches these models.

Our simulations are based on the following moves: a) the community is composed of a fixed number of individuals,  $N$ ; b) at each time step, a randomly chosen individual dies and is replaced; c) the replacement can happen in two distinct ways: i) with probability  $\nu$  a new species is created with abundance 1; ii) with probability  $1 - \nu$  we sample  $m$  individuals at random and compete the first with the second, the winner with the third, etc. The winner of the last competitive bout is chosen for replacing the individual that dies; d) whenever two individuals compete, their probability of winning is encoded in the matrix  $\mathbf{H}$ ; e) new species establish random interactions with the residents (i.e., the corresponding column and row in  $\mathbf{H}$  are randomized).

The transition rates of this model are very similar to the ones obtained for closed systems, and can be written as

$$W_{ij}^{(m)} = x_j \left( (1 - \nu) q_i^{(m)} + \nu \delta_{i,S+1} \right), \quad (\text{S28})$$

where the term  $\nu \delta_{i,S+1}$  is the speciation term and constrain the new species that appears to be a new one. The term  $q_i^{(m)}$  is given by Eq. S8 for the case of equal physiological rates.

Note that for  $m = 1$  we obtain  $q_i^{(1)} = x_i$ , and thus interactions between the species do not play a role, such as in neutral models; for  $m = 2$  a mean-field, deterministic equilibrium exists, but is not attractive; for  $m > 2$  the equilibrium exists and is attractive. We can therefore contrast directly the results obtained for neutral models with those in which species interact in an increasingly higher-order fashion. Note that, for simplicity and to keep the analogy with neutral models, we are implicitly assuming equal fertilities and death rates for all species.

The algorithm is repeated for a sufficiently large number of time steps, and snapshots of the number of extant species and their abundance are taken at intervals that are sufficiently large to

ensure that the snapshots are statistically independent. In order to determine the spacing between the snapshots (i.e., the typical relaxation time scale of the model), we start the system with a single species with abundance  $N$ , and run the simulations until the species goes extinct—at time  $T$ . We then take 1000 snapshots of the system each spaced  $T$  time steps apart.

We ask three questions: a) can we assemble large, diverse communities? b) what is the shape of the species abundance distribution? c) what is the relation between age and abundance of a species?

### S5.1 Assembly and expected number of species

We start the system with a single species, at abundance  $N$  (we choose four different values of  $N$ : 500, 1000, 2500, 5000). We set the speciation rate  $\nu$  to one of five values: 0.0001, 0.0005, 0.001, 0.005, 0.01. Finally, we choose a number of seedlings for each competitive bout,  $m$ : 1 (neutral model), 2, 3, 10. Each simulation is replicated ten times, for a grand total of  $4 \times 5 \times 4 \times 10 = 800$  simulations. For each simulation, we take 100 snapshots spaced as described above to ensure independence.

Neutral theory predicts the average number of species to be<sup>21</sup>, in the limit of large  $N$ ,

$$\langle S \rangle = 1 - N\nu \log(\nu)/(1 - \nu) . \quad (\text{S29})$$

In Fig. S9, we show that the expected number of species, obtained averaging over all the snapshots and replicates, matches closely with what expected under neutrality, irrespective of the number of seedlings  $m$ .

This is an important result, because it means that we can assemble large communities even when species do interact with each other. This is not trivial, given that we showed above that in the deterministic approximation we cannot have an even number of species coexisting, and hence we could not move from one species to three in this setting. We note that there seem to be a small, yet systematic, trend in which sampling more seedlings ( $m > 1$ ) results in elevated number of species at equilibrium. Though this effect seems small and is only apparent for small values of  $\nu$ , it should be investigated further.

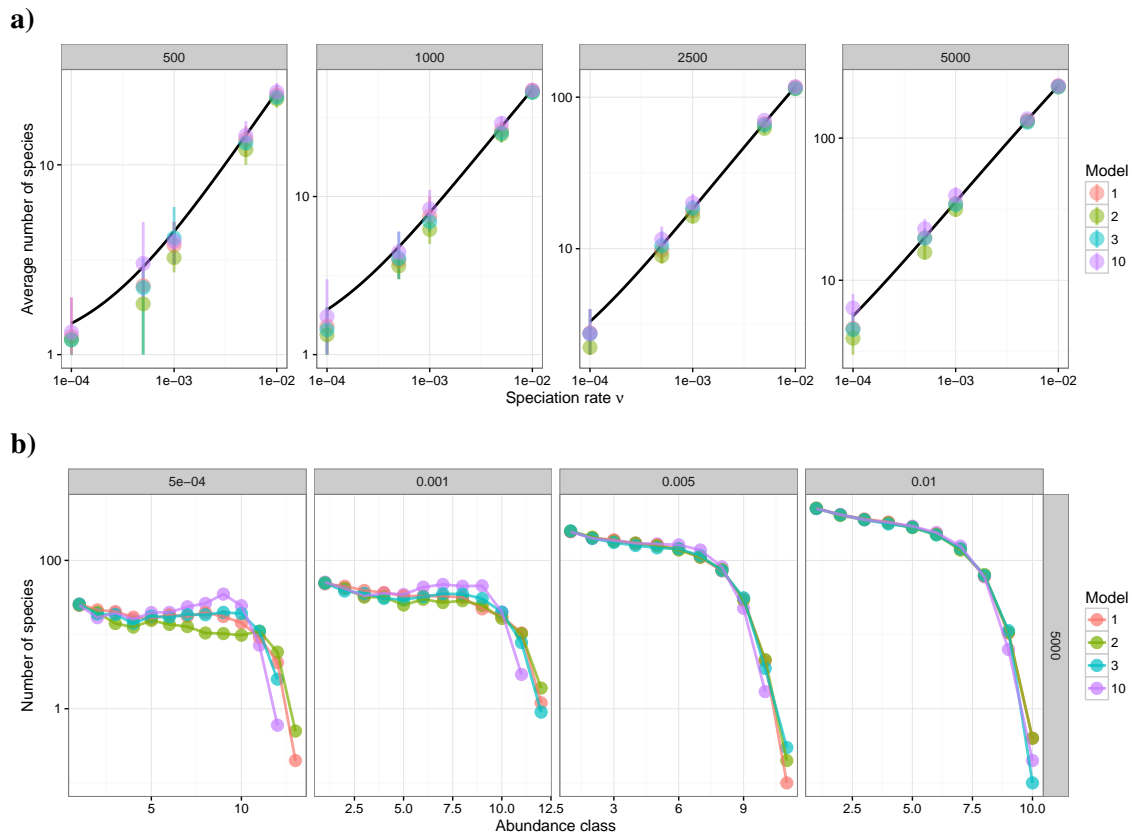
### S5.2 Species-abundance distribution

The ability to reproduce empirically observed Relative Species Abundance (RSA) patterns is one of the main successes of neutral theory. Fig. S9 shows the RSA (averaged over snapshots and replicates) for different values of the speciation rate  $\nu$  and the number of seedlings  $m$ . Surprisingly, neutral theory predictions are very similar to the ones of interacting models, independently of  $m$ . This suggests that the prediction of neutral theory are extremely robust, and are valid even when neutrality is broken. On the other hand, it also implies that RSA are not particularly informative of the dynamical properties of ecological communities<sup>32,33</sup>—RSA is by and large determined by demographic stochasticity that dominates the dynamics of rare species.

### S5.3 Age and abundance

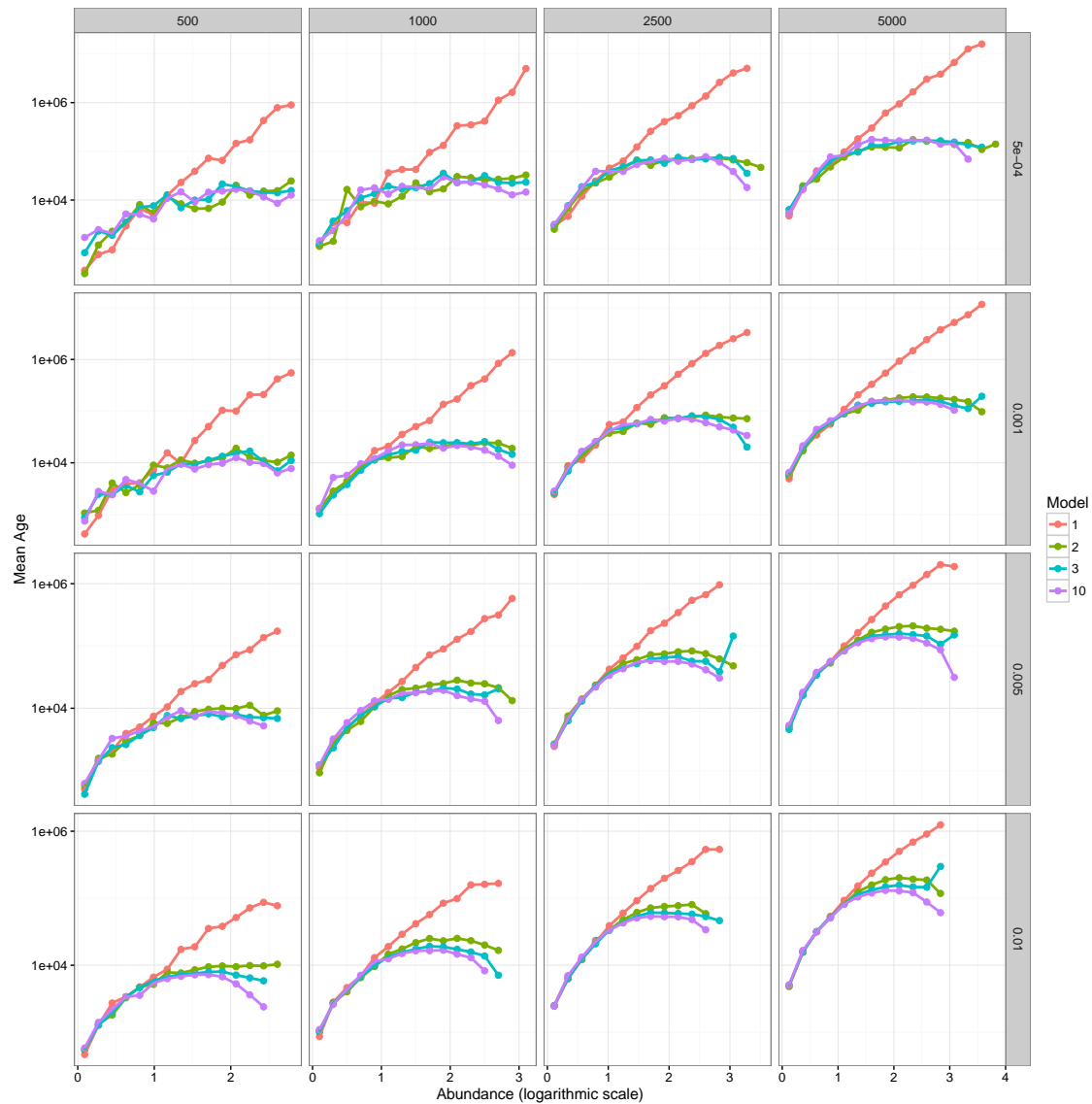
While neutral theory predicts and connects many ecological patterns including spatial and dynamical ones<sup>23</sup>, it fails at predicting quantities on an evolutionary time scale. In particular, neutral models predict a monotonic relationship between species ages and their abundances<sup>28,29</sup>. When applied to tropical rainforests, this relationship would paradoxically imply that the most abundant species would be older than the Earth itself<sup>27,34</sup>.





**Figure S9: Diversity in stochastic simulations with immigration.** Panel a) shows the average number of extant species in the stochastic model with immigration/speciation. The black line represents the expectation  $1 - Nv \log(v)/(1 - v)$  for neutral theory<sup>21</sup>. The error bars enclose 90% of the distribution. Panel b) shows the Relative species abundance, averaged over snapshots and replicates. The RSA is reported using a logarithmic scale (Preston plot)<sup>21</sup>.

We performed an analysis similar to the one shown in Chisholm & O'Dwyer<sup>28</sup>. Figure S10 shows the prediction for age-abundance relationship for different values of population size  $N$ , speciation rate  $v$  and number of seedlings  $m$ . For  $m = 1$  we recover a neutral model and the relation is monotonic (an analytical formula can be found in O'Dwyer & Chisholm<sup>29</sup>). For  $m > 1$ , we obtain that for small abundances, whose dynamics is likely dominated by demographic stochasticity, age and abundance are still positively correlated. For larger abundance, on the other hand, the predictions of neutral models and the ones for models of interacting species strongly diverge. In particular, we obtain that, for  $m > 1$ , the relation reaches a plateau, therefore strongly reducing the predicted age for very abundant species.



**Figure S10: Species age in stochastic simulations with immigration.** Age-abundance relationship averaged over species, snapshots and replicates. For the neutral model ( $m = 1$ , red dots and line) we expect a monotonically increasing relationship<sup>28,29</sup>, while in models with competition this correlation is lost for large abundances.

## References

- [1] Jonasson, J. On the optimal strategy in a random game. *Electronic Communications in Probability* **9**, 132–139 (2004).
- [2] Brandl, F. The distribution of optimal strategies in symmetric zero-sum games. *arXiv preprint arXiv:1611.06845* (2016).
- [3] Maybee, J. S. & Pullman, N. J. Tournament matrices and their generalizations, I. *Linear and Multilinear Algebra* **28**, 57–70 (1990).
- [4] Allesina, S. & Levine, J. M. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences USA* **108**, 5638–5642 (2011).
- [5] Fisher, D. C. & Reeves, R. B. Optimal strategies for random tournament games. *Linear Algebra and its Applications* **217**, 83–85 (1995).
- [6] Fisher, D. C. & Ryan, J. Optimal strategies for a generalized “scissors, paper, and stone” game. *American Mathematical Monthly* **99**, 935–942 (1992).
- [7] Buss, L. & Jackson, J. Competitive networks: nontransitive competitive relationships in cryptic coral reef environments. *The American Naturalist* 223–234 (1979).
- [8] McCoy, S. & Pfister, C. Historical comparisons reveal altered competitive interactions in a guild of crustose coralline algae. *Ecology Letters* **17**, 475–483 (2014).
- [9] Benincà, E., Ballantine, B., Ellner, S. P. & Huisman, J. Species fluctuations sustained by a cyclic succession at the edge of chaos. *Proceedings of the National Academy of Sciences USA* **112**, 6389–6394 (2015).
- [10] Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418**, 171–174 (2002).
- [11] Czárán, T. L., Hoekstra, R. F. & Pagie, L. Chemical warfare between microbes promotes biodiversity. *Proceedings of the National Academy of Sciences USA* **99**, 786–790 (2002).
- [12] Keddy, P. A. & Shipley, B. Competitive hierarchies in herbaceous plant communities. *Oikos* 234–241 (1989).
- [13] Goldberg, D. E. Competitive ability: definitions, contingency and correlated traits. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **351**, 1377–1385 (1996).
- [14] Soliveres, S. *et al.* Intransitive competition is widespread in plant communities and maintains their species richness. *Ecology Letters* **18**, 790–798 (2015).
- [15] Biancalani, T., Dyson, L. & McKane, A. J. Noise-induced bistable states and their mean switching time in foraging colonies. *Physical Review Letters* **112**, 038101 (2014).
- [16] Capitán, J. A., Cuenda, S. & Alonso, D. How similar can co-occurring species be in the presence of competition and ecological drift? *Journal of The Royal Society Interface* **12**, 20150604 (2015).

- [17] Reichenbach, T., Mobilia, M. & Frey, E. Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model. *Physical Review E* **74**, 051907 (2006).
- [18] Caswell, H. Community structure: A neutral model analysis. *Ecological Monographs* **46**, 327–264 (1976).
- [19] Hubbell, S. P. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16**, S9–S21 (1997).
- [20] Abrams, P. A. A world without competition. *Nature* **412**, 858–859 (2001).
- [21] Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*, vol. 32 (Princeton University Press, 2001).
- [22] McKane, A., Alonso, D. & Solé, R. V. Mean-field stochastic theory for species-rich assembled communities. *Physical Review E* **62**, 8466 (2000).
- [23] Azaele, S. *et al.* Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics* **88**, 035003 (2016).
- [24] Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037 (2003).
- [25] Alonso, D., Etienne, R. S. & McKane, A. J. The merits of neutral theory. *Trends in Ecology & Evolution* **21**, 451–457 (2006).
- [26] Chisholm, R. A. *et al.* Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecology Letters* **17**, 855–865 (2014).
- [27] Nee, S. The neutral theory of biodiversity: Do the numbers add up? *Functional Ecology* **19**, 173–176 (2005).
- [28] O’Dwyer, J. P. & Chisholm, R. A mean field model for competition: from neutral ecology to the Red Queen. *Ecology Letters* **17**, 961–969 (2014).
- [29] Chisholm, R. A. & O’Dwyer, J. P. Species ages in neutral biodiversity models. *Theoretical Population Biology* **93**, 85–94 (2014).
- [30] Solé, R. V., Alonso, D. & McKane, A. Scaling in a network model of a multispecies ecosystem. *Physica A: Statistical Mechanics and its Applications* **286**, 337–344 (2000).
- [31] Solé, R. V., Alonso, D. & McKane, A. Self-organized instability in complex ecosystems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **357**, 667–681 (2002).
- [32] Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: theoretical consequences for patterns of diversity. *The American Naturalist* **159**, 1–23 (2002).
- [33] Purves, D. W. & Pacala, S. W. Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. *Biotic Interactions in the Tropics* 107–138 (2005).
- [34] Ricklefs, R. E. A comment on Hubbell’s zero-sum ecological drift model. *Oikos* **100**, 185–192 (2003).